

OM of: US-08-973-363-6 to: EST.\* out\_format: pfs

Date: Aug 3, 2002 3:03 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

MODE=fasta\_p2n model -dev=11h  
-o=/gen2/USPTO/spool/US08973363/finat\_01082002\_080049\_18570/app\_query.fasta\_1.638  
-b=EST -qmt=fastap -suffix=p2n.rst -GAPOP=12.000 -GAPEXT=4.500  
MINMATCH=0.100 -LOOPEL=0.000 -LOPEXT=0.000 -GAPOP=6.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELTA=6.000  
-DELTA=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pts -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US08973363\_@CGN1\_1.2938  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-08-973-363-6  
Query length: 41  
Database: EST.\*  
Database sequences: 13736207  
Database length: 1841457050  
Search time (sec): 4390.400000

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est1.BB155356	+	208.00	528.75	3.3e-20	619	BI155356 BB155356 RIKEN full-16
gb_est1.BB161065	+	208.00	528.11	3.6e-20	619	BB161065 BB161065 RIKEN full-16
gb_est1.BB834922	+	203.00	518.98	1.2e-19	446	BB834922 BB834922 RIKEN full-16
gb_est1.BB830730	+	193.00	493.12	3.2e-18	438	BB830730 BB830730 RIKEN full-16
gb_est1.AL601246	+	174.00	443.82	1.8e-15	430	AL601246 DKFZ03131040.r1.313
gb_est1.A1690775	+	174.00	441.42	2.4e-15	547	AL690775 wms5f11.x1 NCL CGAP v1
gb_est1.AU125712	+	174.00	436.86	4.4e-15	866	AU125712 AU125712 NM2200-102
gb_est1.BE951133	+	174.00	435.15	5.4e-15	1028	BE951133 60141606061 NIH-MGC-7
gb_est1.AL645934	+	172.00	434.58	5.3e-15	645	AL645934 AL645934 XCC-neurula
gb_est1.BF239967	+	147.00	367.07	3.4e-11	821	BF239967 601905170F1 NIH-MGC-54
gb_gss.CMS04DNG	-	139.50	356.35	1.3e-10	359	AL286251 Tetradodon nigroviridis
gb_est1.BM936787	-	139.00	348.02	3.9e-10	686	BM936787 QV3-BM0047-230200-102
gb_gss.CMS07058	+	121.00	296.25	3.0e-07	1122	AW997058 QV3-BM0047-150400-132
gb_est1.B1844947	-	119.00	297.70	2.5e-07	532	AL352864 Tetradodon nigroviridis
gb_est1.BM185888	-	119.00	297.13	2.5e-07	574	BM185888 fql7e05.x1 zebrafish a
gb_est1.BM070535	-	119.00	297.63	2.6e-07	578	BM070535 f099h05.x1 zebrafish a
gb_est1.BM074625	-	119.00	297.13	2.6e-07	608	BM074625 f099h05.x1 zebrafish a
gb_est1.BM185005	-	119.00	297.13	2.6e-07	608	BM185005 f16c08.x2 zebrafish a
gb_gss.CMS04J51	+	119.00	292.26	3.4e-07	734	AL133535 Tetradodon nigroviridis
gb_gss.CMS04X68	+	119.00	292.98	4.5e-07	923	AL133535 Tetradodon nigroviridis
gb_gss.CMS05KJ7	+	119.00	292.56	4.8e-07	963	AL171305 Tetradodon nigroviridis
gb_gss.CMS05NDJ	+	119.00	292.04	5.1e-07	1015	AL141950 Tetradodon nigroviridis
gb_gss.CMS05N3J	+	119.00	291.70	5.3e-07	1050	AL344672 Tetradodon nigroviridis
gb_gss.CMS05O6W	+	119.00	291.66	5.3e-07	1054	AL344672 Tetradodon nigroviridis
gb_gss.CMS0510P	+	119.00	291.48	5.5e-07	1074	AL317122 Tetradodon nigroviridis
gb_gss.CMS051C3	+	119.00	291.23	5.6e-07	1096	AL316668 Tetradodon nigroviridis
gb_gss.CMS04U0A	+	118.00	301.45	1.5e-07	1101	AL308107 Tetradodon nigroviridis
gb_est1.A439703	-	118.00	300.58	1.7e-07	300	NA9703 yv23603.s1 Soares fel1
gb_est1.A305759	-	118.00	299.57	1.9e-07	334	AA305759 EST176834 Colon carcin
gb_est1.A1915135	-	118.00	297.83	2.4e-07	366	AI1915135 w150e08.x1 NCL CGAP.CC
gb_est1.A1648356	-	118.00	297.61	2.5e-07	436	AI648356 t255b06.x1 NCL CGAP.CC
gb_est1.A53364	-	118.00	295.60	3.2e-07	446	H53364 y08b11.s1 Soares fel1
gb_est1.BG757172	-	118.00	293.61	4.2e-07	667	BG757172 602710541F1 NIH-MGC-48
gb_est1.AW364689	-	118.00	292.75	4.6e-07	724	AW364689 PM3-PM0037-231299-001
gb_gss.CMS05D5J	+	118.00	289.92	6.7e-07	967	AL332882 Tetradodon nigroviridis
gb_gss.CMS02B5S	+	118.00	289.30	7.2e-07	967	AL189793 Tetradodon nigroviridis
gb_gss.CMS05G5M	-	118.00	289.30	7.2e-07	1039	AL336703 Tetradodon nigroviridis
gb_gss.CMS04VNV	-	118.00	289.23	7.3e-07	1036	AL309919 Tetradodon nigroviridis

gb\_gss.CMS04VNV - 118.00 289.23 7.3e-07 1036 AL309919 Tetradodon nigroviridis

gb\_gss.CMS04VNV - 118.00 289.13 7.4e-07 1047 AL292833 Tetradodon nigroviridis

gb\_gss.CMS02B5S - 118.00 289.09 7.4e-07 1051 AL192743 Tetradodon nigroviridis

gb\_gss.CMS05R92 - 118.00 289.04 7.5e-07 1056 AL350288 Tetradodon nigroviridis

gb\_gss.CMS05H1G - 118.00 288.83 7.7e-07 1079 AL357633 Tetradodon nigroviridis

seq\_name: gb\_est1:BB155356

seq\_documentation\_block:

LOCUS BB155356 619 bp mRNA linear EST 18-OCT-2001

DEFINITION musculus cDNA clone A130024L16.3 similar to L10410 Mouse DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION BB155356

VERSION BB155356.2 GI:16268254

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,B., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. (2001) Unpublished (2001)

TITLE On Jun 29, 2000 this sequence version replaced gi:8811286.

JOURNAL Contact: Yoshitake Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute, The Institute of Physical and Chemical Research (RIKEN) Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsr.riken.go.jp, url: http://genome.gsc.riken.go.jp/

Arakawa,T., Carninci,P., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Waghi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matsumoto,K., Yoneda,S., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1..619

organism="Mus musculus"

/db\_xref="taxon:10090"

FEATURES

source







genes. Genome Res. . 10 (10), 1617-1630 (2000)  
Wag1.K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Machikita,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichipillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

FEATURES  
source location/Qualifiers  
1..438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone.lib="G930013K04"  
/clone.lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B)"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT 138 a 96 c 108 g 96 t

ORIGIN

alignment\_scores:  
Quality: 193.00 Length: 42  
Ratio: 4.707 Gaps: 1  
Percent Similarity: 97.619 Percent Identity: 95.238

alignment\_block:  
US-08-973-363-6 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

1 lleleuproaspappproaspplyslysproglnalalysglnleuglnth 17  
|||||  
300 ATTCTTCACATGATGCTCGATTAATAAACCAACCAACCAACGACTTACGAC 349  
17 fargalaspptyrleuileuileusleuuserargaspplenualalysarg 33  
|||||  
350 CCGTCGACGACTACCTCCTCAACAGCTTACGACGAGATCTGCACAAAAGA 399  
34 GluAlaGlnArgLeuGlyGlyAla 41  
|||||  
400 GAGGCTCAGAGACTTGTGTGCG 423

seq\_name: gb\_est1:AL601246

seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-ANG-2001  
DEFINITION DKFZP313J1040.r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKEP313J1040 5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Weil,B. and Wiemann  
S.  
TITLE EST (Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Weil,B. and Wiemann  
S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloecker H  
MIRS  
Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
source location/Qualifiers  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone.lib="DKFZP313J1040"  
/clone.lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pTriplex2; site\_1: SfiIA; site\_2: SfiIB;  
cDNA-collection"

BASE COUNT 163 a 81 c 86 g 100 t

ORIGIN

alignment\_scores:  
Quality: 174.00 Length: 41  
Ratio: 4.579 Gaps: 1  
Percent Similarity: 92.683 Percent Identity: 90.244

alignment\_block:  
US-08-973-363-6 x AL601246 ..

Align seg 1/1 to: AL601246 from: 1 to: 430

1 lleleuproaspappproaspplyslysproglnalalysglnleuglnth 17  
|||||  
218 ATTCTTCACATGATGCTCGATTAATAAACCAACCAACCAACGACTTACGAC 267  
17 fargalaspptyrleuileuileusleuuserargaspplenualalysarg 34  
|||||  
268 CCGTCGACGACTACCTCCTCAACATTAATGTAAGAGATCTGCACAAAAGA 317  
34 LuAlaGlnArgLeuGlyGlyAla 41  
|||||  
318 AAGCT.....CTTTCGTGTCGCG 334

seq\_name: gb\_est1:A1890775

seq\_documentation\_block:  
LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wms95f11.x1 NCI\_CGAP.Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1 ; , mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/brp/image/image.html  
 Insert length: 1924 Std Error: 0.00  
 Seq primer: 400P from Glbco  
 High quality sequence stop: 418.

## FEATURES

source  
 Location/Qualifiers  
 1..547  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="2443725"  
 /clone\_lib="NC1,CGAP,Ut2"  
 /tissue\_type="moderately-differentiated endometrial  
 adenocarcinoma, 3 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-Sport6; Site\_1: Salt;  
 Site\_2: Not; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.85 kb. Life Technologies catalog #:  
 11539-012"  
 BASE COUNT 114 a 118 c 85 g 230 t  
 ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
 Ratio: 4.579 Gaps: 1  
 Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

```

1 11leuProaspapProaspyslyspProglAlaIysGlnIeuGlnth 17
|||||
378 ATTCTTCAGATGATCCCGATMAAACACACACACACACTTCGACAC 329
|||||
17 TATGAlAspTyrIeuIleIysIeuIeuSerArgAspIeuAlaIysArg 34
|||||
328 CCGTGAAGACTACCTCATCAATCTTAGAGAGATCTGCAGAAAAAG 279
|||||
34 TAAAGAlnArgIeuCysGlyAla 41
|||||
278 AAGCT.....CTTCTGTGCG 262

```

seq\_name: gb\_est1:AU125712

## seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
 DEFINITION AU125712 Homo sapiens CDNA clone NT2RM4002061 5', mRNA  
 sequence.

ACCESSION AU125712 GI:10950428  
 VERSION AU125712.1  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 866)  
 AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Iisogi,T.

## TITLE

HRT human CDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,  
 Y., Sugano,S., Iisogi,T.)

## JOURNAL

COMMENT Unpublished (2000)  
 Contact: Takao Iisogi  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp

HRT human CDNA project: 5'- & 3'-end one pass sequencing: Helix

Research Institute, CDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers  
 1..866

## FEATURES

source  
 Location/Qualifiers  
 1..866  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="NT2RM4002061"  
 /clone\_lib="NT2RM4"  
 /cell\_type="teratocarcinoma"  
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
 precursor cells"  
 BASE COUNT 312 a 149 c 196 g 207 t 2 others  
 ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
 Ratio: 4.579 Gaps: 1  
 Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AU125712 ..  
 Align seg 1/1 to: AU125712 from: 1 to: 866

```

1 11leuProaspapProaspyslyspProglAlaIysGlnIeuGlnth 17
|||||
450 ATTCTTCAGATGATCCCGATMAAACACACACACACACTTCGACAC 499
|||||
17 TATGAlAspTyrIeuIleIysIeuIeuSerArgAspIeuAlaIysArg 34
|||||
500 CCGTGAAGACTACCTCATCAATCTTAGAGAGATCTGCAGAAAAAG 549
|||||
34 TAAAGAlnArgIeuCysGlyAla 41
|||||
550 AAGCT.....CTTCTGTGCG 566

```

seq\_name: gb\_est2:BE895133

## seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
 DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens CDNA clone IMAGE:3921087 5',  
 mRNA sequence.

ACCESSION BE895133 GI:10358221  
 VERSION BE895133.1  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1028)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: ceasbs-remail.nih.gov  
 Tissue Procurement: ATCC/CDTD/DTF

CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Invitrogen, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: ILAM9753 row: h column: 16  
 High quality sequence stop: 488.

## FEATURES

source  
 Location/Qualifiers  
 1..1028  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3921087"  
 /clone\_lib="NIH\_MGC\_72"

```
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: PCMV-SPOrt6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT      387 a      205 c      238 g      198 t
ORIGIN

alignment_scores:
      Quality: 174.00      Length: 41
      Ratio: 4.579      Gaps: 1
Percent Similarity: 92.683      Percent Identity: 90.244

alignment_block:
US-08-973-363-6 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 11leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
131 ATTCTTCCAGATGATCCCGATTAACACACACAAACAGCTTGACAGAC 180
|||||
17 TATGAlAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArg 34
|||||
181 CCGTGACAGACTACTCATCAATAGTAGAGATCTTGCAAAAAAAG 230
|||||
34 1uAlaGlnArgLeuCysGlyAla 41
|||||
231 AACCT.....CTTTCGTGTCG 247

seq_name: gb_est1:AL659353

seq_documentation_block:
LOCUS      AL659353      593 bp      mRNA      linear      EST 13-DEC-2001
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5',
            mRNA sequence.
ACCESSION  AL659353
VERSION    AL659353.1 GI:17672995
KEYWORDS  EST.
SOURCE     western clawed frog.
ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
            Xenopodinae; Silurana.
            1 (bases 1 to 593)
REFERENCE  1 Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (10_2001)
            Unpublished (2001)
            Contact: Huckle E
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: tropesanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TNeu045e20.sp6
            Sequencing primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES
            source
            1..593
            /organism="Silurana tropicalis"
            /db_xref="taxon:8364"
            /clone="TNeu045e20"
            /clone_lib="XGC-neurula"
            /dev_stage="neurula"
            /lab_host="Escherichia coli DH10B"
            /note="Vector: PCS107; Site: 1: EcoRI; Site: 2: NotI; cDNA
            was oligo dt primed from 5ug of poly A+ RNA from neurula.
            EcoRI-NotI cut cDNA was then ligated into PCS107 with
            EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      119 a      134 c      114 g      226 t
ORIGIN
```

```
alignment_scores:
      Quality: 172.00      Length: 38
      Ratio: 4.649      Gaps: 0
Percent Similarity: 97.368      Percent Identity: 86.842

alignment_block:
US-08-973-363-6 x AL659353/rev ..

Align seg 1/1 to reverse of: AL659353 from: 1 to: 593

1 11leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
346 ATTTCACAGATGATCCAGATAGAGAGCCCAAGCAGAGCTACAGAC 297
|||||
17 TATGAlAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArg 34
|||||
296 CAGAGCTGACTACTCCTCAATTAACCTTCATTAAGATCTGCTAGCAAG 247
|||||
34 1uAlaGlnArgLeu 38
|||||
246 AAGCACAAAGACTT 233

seq_name: gb_est1:AL644594

seq_documentation_block:
LOCUS      AL644594      645 bp      mRNA      linear      EST 07-NOV-2001
DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA
            sequence.
ACCESSION  AL644594
VERSION    AL644594.1 GI:16796719
KEYWORDS  EST.
SOURCE     western clawed frog.
ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
            Xenopodinae; Silurana.
            1 (bases 1 to 645)
REFERENCE  1 Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (10_2001)
            Unpublished (2001)
            Contact: Huckle E
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: tropesanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: L1E1d12.pic
            Sequencing primer: PIC
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
FEATURES
            source
            1..645
            /organism="Silurana tropicalis"
            /db_xref="taxon:8364"
            /clone="L1E1d12"
            /clone_lib="XGC-egg"
            /dev_stage="egg"
            /lab_host="Escherichia coli XL1-blue"
            /note="Vector: PCS107; Site: 1: EcoRI; Site: 2: NotI; cDNA
            was oligo dt primed from 5ug of poly A+ RNA from egg.
            EcoRI-NotI cut cDNA was then ligated into PCS107 with
            EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      222 a      125 c      156 g      141 t      1 others
ORIGIN

alignment_scores:
      Quality: 172.00      Length: 38
      Ratio: 4.649      Gaps: 0
Percent Similarity: 97.368      Percent Identity: 86.842

alignment_block:
```

**alignment\_block:**

```
alignment_block:
```

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

```
2 LeuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTrp 18
|||||
674 CTGCGGATATCTCTGATTAAGACCTCAGCCCAACAGTTACAGCAG 625
18 gAlaAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArgGlna 35
|||||
624 AGCCDACTACCTCATCAGCTGCTGACGACCAAGCACTGCGACGAAAGAAG 575
35 lacGlnArgLeuGysGly 40
|||||
574 CCCACAGACGAGSGGGC 558
```

seq\_name: gb\_est1:AW996787

seq\_documentation\_block:

LOCUS AW996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW996787  
VERSION AW996787.1 GI:8257021  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 337)

REFERENCE  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230  
200-102-d03&t3=2000-02-23&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 337.

FEATURES  
source

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast.normal; Vector: puc18; Site.1: SmaI;  
Site.2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 83 a 59 c 72 g 123 t  
ORIGIN

alignment\_scores: Quality: 139.50 Length: 36  
Ratio: 4.103 Gaps: 2  
Percent Similarity: 94.444 Percent Identity: 88.889

alignment\_block:

US-08-973-363-6 x AW996787/rev ..

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```
1 lLeuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTrh 17
|||||
105 ATTCTTCAGATGATCCCGATTAAGACCAACCAACAGTTGACAGAC 56
17 rArgAlaAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArg 33
|||||
34 GlnAla 35
|||||
8 GAAGCT 3
```

seq\_name: gb\_est1:AW997058

seq\_documentation\_block:

LOCUS AW997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW997058  
VERSION AW997058.1 GI:8257292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 686)

REFERENCE  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150  
400-152-c03&t3=2000-04-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 678.

FEATURES  
source

1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast.normal; Vector: puc18; Site.1: SmaI;  
Site.2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 165 a 154 c 126 g 241 t  
ORIGIN

## alignment\_scores:

Quality: 139.00 Length: 35  
Ratio: 3.971 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 94.286

## alignment\_block:

US-08-973-363-6 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

```

1 11leuPProaspAspProaspLysLysPProGlnAlaLysGlnLeuGlnTh 17
|||||
127 ATTCTTCAGATGATCCGATTA.AAACCAAGCAAAACGTTGCAGAC 79
|||||
17 rArgAlaAspTyrLeu11eLysLeuLeuSerArgAspLeuAlaLysArg 34
|||||
78 CCGTCAGACTGATCTCATCAATTAATTACTAGAGACTTGCACAAAAGAA 29
|||||
34 luAla 35
|||||
28 GCTCT 24

```

seq\_name: gb\_gss:CNS05T9J

seq\_documentation\_block:

LOCUS CNS05T9J 1122 bp DNA linear GSS 26-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
042M09 of library C from Tetraodon nigroviridis, genomic survey  
sequence.  
AL352864.1 GI:8246657  
VERSION AL352864  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 1122)  
Roest-Crollius H., Jallion O., Dasilva C., Fizames C., Fisher C.,  
Bonneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
Weissenbach J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 1122)  
Roest-Crollius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,  
Bernot A., Fizames C., Wincker P., Brothier P., Quetier F.,  
Saurin W. and Weissenbach J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 1122)  
Genoscope.

JOURNAL Direct Submission  
REFERENCE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
AUTHORS This sequence is a single read and was generated as part of a large  
JOURNAL scale clone-end sequencing project of the Tetraodon nigroviridis  
COMMENT genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1..1122  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_11b="C"  
/clone\_11d="C"  
/note="Genoscope sequence ID : C0BC042AG05C1-end : T7"

## FEATURES

BASE COUNT 303 a 260 c 301 g 246 t 12 others  
ORIGIN

alignment\_scores: Quality: 121.00 Length: 29

Ratio: 4.172 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.862

## alignment\_block:

US-08-973-363-6 x CNS05T9J ..

Align seg 1/1 to: CNS05T9J from: 1 to: 1122

```

1 11leuPProaspAspProaspLysLysPProGlnAlaLysGlnLeuGlnTh 17
|||||
835 ATTCTTCAGATGATCCGATTA.AAACCAAGCAAAACGTTGCAGAC 884
|||||
17 rArgAlaAspTyrLeu11eLysLeuLeuSerArgAsp 29
|||||
885 CAGAGCTGAGTATCTCTCTCAAGCTGCTGAAAAAGAC 921
|||||

```



...



```
source 1. .6608
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2289 a 1207 c 1459 g 1653 t
ORIGIN

alignment_scores:
  Quality: 205.00 Length: 41
  Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x A58691 ..

Align seg 1/1 to: A58691 from: 1 to: 6608

1 lileuproaspaspProaspLysProGlnAlaLysGlnLeuGlnIth 17
|||||
4080 ATTTTACCTGATGATCCAGACAAAGAACCCAGCAAGCCTTGCAAGAAG 4129
17 fargAlaAspTyrLeuIleuLysLeuAsnLysAspLeuAlaArgLysG 34
|||||
4130 CCGTGCAGACTACTCTATTAAATTACTGAATTAAGACCTTGCAAGAAG 4179
34 luAlaGlnArgLeuAlaGlyAla 41
|||||
4180 AAGCACAAGAGCTTGCTGTGCA 4202

seq_name: gb_ov:AF004397

seq_documentation_block:
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION Gallus chromo-helicase-DNA-binding on the z chromosome
protein, variant with hydrophilic domain. (CHD-2) mRNA, complete
cds
ACCESSION AF004397
VERSION AF004397.1 GI:2501845
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 6872)
REFERENCE
AUTHORS Griffiths,R. and Korn,R.M.
TITLE A CHD1 gene is z chromosome linked in the chicken Gallus domesticus
JOURNAL Gene 197 (1-2), 225-229 (1997)
MEDLINE 97473516
REFERENCE
AUTHORS Griffiths,R. and Korn,R.M.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,
Glasgow G12 8QQ, UK

FEATURES
source
location/Qualifiers
1. .6872
/organism="Gallus gallus"
/db_xref="taxon:9031"
1. .6872
/gene="CHD-2"
228. .5654
/gene="CHD-2"
/function="role in chromatin architecture"
/note="CHD protein with hydrophilic domain"
/codon_start=1
/product="chromo-helicase-DNA-binding on the z chromosome
protein"
/protein_id="AAC60282.1"
/db_xref="GI:2501846"
/translation="MNGHSDSESVNNSGESSRSDDGASAGSGSGSSGSSSSGSS
SQSGSDSESSGSGSSESDTRREKQVAKPKPADGSEFMKSPBITAYQSAV
LKQOQOQKAASDSGSESDSSADSSSEFKKKHKHDKEDQMGSGSGSVGTGS
DSEAEDEKSSCESESDYEPKKNVKSRRPSRIKPSGKKSTGQKRQDLSSEEE
```

```
DDDEDYDKRGRQATVNVSYKEAEETKTDSDLLVEGDEVDPQTEDEFEETIEKFM
SRIGRGKATGASTTIIYAAEDADPNPAEKESELEIGIYLILKMGMSHINHWTEET
LKQONVKGNKLDNKKKDOETKRWLKNASPDVYVYNOQOELTDLKKOYIVRETI
AHSNOKSAAGPYDYCKKNOGLPYSECSMDGALLAKKQARIDEYFENOSKTFPKD
CYLKKRPFVALKKOPSTIGHESLELDYOLNGLNASHWCNGNSCTLADENGLG
KTIGTISPLNTLFEHDLXGPPFLRPLSTLTSWQREIQTWAPQNAVAYVLTIDSRN
MIRHEMMHPQTRLRKFNILITVEILILDKSFLGLNMAFGVDEHRLKNDLSLY
RTLIDPKSNHRLITGTPLONSLKEMLHIFIMPEKSWMEDFEHKGKRGYAS
LHKELEPFLRRVKKDVESLPAKVEOLRMEASALOKOYKWIITRYN KALSKSGK
STSGFLNIMELKKCCNHCYLKPPDDNPEYFKKQALOHILRSSGKLTLLKLLRLR
EKGNYLIFSQYVRMLDILAEYLKTRQPPQNLDSIKGELRKALDHPNMGSEDFC
FLSTRAGGLGINLASADTVLVEDSDMNPQNDLQADARHRIQKQVNIYKLVYGS
VEEDILIERAKKMMVLDHLYTQMDDTGKTVLHTGSPSSPTPNKEELSAILKEGAE
LEKEPEGEEOPEQMDIDELIKRAETRENEPGLVGEDLLSQFYVANSNDEDDIE
LEPERNSRWMEIIPESORRIIEEEROKELEIYMLPRMCAKOISPNSEGRSR
SRYSQSDSTTEKRPKRRPRPTIRENTKGSQSGORRAGRLGKGPFRISGV
AVARDELVDKSETDLRIGELVHNGCTIKALDNSSGOBRAGRLGKGPFRISGV
OVNARVLISHEELAPLHKSTPSDEERKRYVIPHOTAAHEDIDWGEDSNLLVGI
VEYGYGSMEMIKMDPDLSTOKILPDDPPKKQAOLOTRADYLYLKLNLKDLAREAQ
FLAGAGNSKRRRTRNKKNNKASKIKEETSDSPSEKSEDEDEEDNDKEIYSVK
HLAKKITKEKENEKPEPDIGIKKEAEERETKEKNEBELKREKEDKKLEKD
NKEKRENNKYESTOKEKYEKPEKVENKSEKSKTPLDTPYHITATSPPLISE
ESEEHLQKTFSTYCKERMRPVKALYQDLRPEKGLSEROLETTRQCLIKIGHTTECL
KEYTPEQIKQMRKNMLFVSKETFEFDARKLRLYKHA1IKRQESQOHNDQNTSSNVN
THVINRPDYERLKEETTNDDSSRDSYSDRHLISYHDKHDKHODDAYKKSDDRPPY
SAFNSKDHDRMDHYKODSRYSDSKHRLDHRSDRHSNLEGMILKDSRSDHRS
SDRHSKDHRSSTSEYSHKSSRDYRHSQWQMDHRASSGSPRPLDQSPYCSRPPLG
HSPREHSDHKSTPEHTWSSKRT"
4341. .4604
/gene="CHD-2"
/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t
ORIGIN

alignment_scores:
  Quality: 205.00 Length: 41
  Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 lileuproaspaspProaspLysProGlnAlaLysGlnLeuGlnIth 17
|||||
4080 ATTTTACCTGATGATCCAGACAAAGAACCCAGCAAGCCTTGCAAGAAG 4129
17 fargAlaAspTyrLeuIleuLysLeuAsnLysAspLeuAlaArgLysG 34
|||||
4130 CCGTGCAGACTACTCTATTAAATTACTGAATTAAGACCTTGCAAGAAG 4179
34 luAlaGlnArgLeuAlaGlyAla 41
|||||
4180 AAGCACAAGAGCTTGCTGTGCA 4202

seq_name: gb_pat:A58684

seq_documentation_block:
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from patent WO9639505.
ACCESSION A58684
VERSION A58684.1 GI:3714247
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
BIRDS
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;
```

```

COMMENT      ISIS INNOVATION (GB)
FEATURES     Other publication AU 5906996 961224.
              Location/Qualifiers
                source          1..153
                                /organism="unidentified"
                                /db_xref="taxon:32644"
BASE COUNT   58 a         40 c        31 g        24 t
ORIGIN
Alignment_scores:
               Quality: 190.00      Length: 51
               Ratio:    4.634       Gaps:    1
Percent Similarity: 80.392      Percent Identity: 80.392

seq_name: gb_pat:A58685
Align seq 1/1 to: A58684 from: 1 to: 153
US-08-973-363-7 x A58684 ..

Alignment_block:
US-08-973-363-7 x A58685 ..

1 TlaleuProaspasPproasp.....
|||||
1 ATTTACTGATGATTCACAGCAAGAAACCCAGCGAAGCAGCTACAGAC 50
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||
51 CAAGAACCAGGCCAGCAAGCAGCTACAGACCCGTGCAGACTACTCATTA 100
24 ystleuenuasmLysAspleuAlaArgLysGlnAlaArgLeuAlaGly 40
|||||
101 AATTACTGATTAAGACCTTGCAAGAAAGAACCAAGCAGCTTGCGGT 150
41 Ala 41
|||
151 GCA 153

seq_name: gb_pat:A58685
seq_documentation_block:
LOCUS      A58685             153 bp      DNA      linear      PAT 06-MAR-1998
DEFINITION Sequence 4 from Patent WO9639505.
VERSION     A58685
KEYWORDS    A58685.1 GI:3714248
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tiwari,B.
AVIAN GHF GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 4 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES     Location/Qualifiers
               source          1..153
                                /organism="unidentified"
                                /db_xref="taxon:32644"
BASE COUNT   56 a         36 c        31 g        30 t
ORIGIN
Alignment_scores:
               Quality: 190.00      Length: 51
               Ratio:    4.634       Gaps:    1
Percent Similarity: 80.392      Percent Identity: 80.392

seq_name: gb_ro:MUSCHD1X
Align seq 1/1 to: A58684 from: 1 to: 153
US-08-973-363-7 x A58684 ..

Alignment_block:
US-08-973-363-7 x A58685 ..

1 TlaleuProaspasPproasp.....
|||||
1 ATTTACTGATGATTCACAGCAAGAAACCCAGCGTACAGCTACAGAC 50
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||
51 CAAGAACCAGGCCAGCAAGCAGCTACAGACCCGTGCAGACTACTCATTA 100
24 ystleuenuasmLysAspleuAlaArgLysGlnAlaArgLeuAlaGly 40
|||||
101 AATTACTGATTAAGACCTTGCAAGAAAGAACCAAGCAGCTTGCGGT 150
41 Ala 41
|||
151 GCA 153

seq_name: gb_ro:MUSCHD1X
seq_documentation_block:
LOCUS      MUSCHD1X           5349 bp      mRNA      linear      ROD 19-SEP-1996
DEFINITION Mouse X66028 mouse X66028
ACCESSION   L10410.1
VERSION     L10410.1 GI:455014
KEYWORDS    DNA binding protein.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 5349)
AUTHORS     Delmas,V., Stokes,D.G. and Perry,R.P.
TITLE       A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)
MEDLINE     93211972
REFERENCE   2 (bases 1 to 5349)
AUTHORS     Perry,R.P.
TITLE       Direct Submission
JOURNAL     Submitted (08 APR 1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA
COMMENT     On Feb 16, 1994 this sequence version replaced gi:293322.
FEATURES     Location/Qualifiers
               source          1..5349
                                /organism="Mus musculus"
                                /db_xref="taxon:10090"
                                /cell_type="plasmacytoma"
               gene            1..5349
                                /gene="CHD-1"
                                /protein_id="AAB08486.1"
               CDS             171..5306
                                /gene="CHD-1"
                                /product="DNA-binding protein"
                                /protein_id="GI:455015"
               translation     "MNGSHDESVRNGSGESSQSDGCSASGSSSSSSSDGSS
SQSSSSDSGSDSGSSESDPTRENNKVOAPPVDAEFKMSPSILAYVRAMML
RKPOQAQQOQRASNSGEEDSSSEDSSGGAKRKINDMDMGSGSPQGLAQ
SDSESEERDKSCDGTESDYEKRRNKRVSRRPNRSKNKKLLGOKROIDSDE
DDBDYNDKRSRRQATVANYSYKEDBEKMTDSDLEVGCVQDPDEFEFTIR
DCVGRKGATGATTTTIVAVADGDNPAGFNERNPEPDIOYLICKRGSHIHNTWTEET
TIKOONVRGMKLIDNYKKCKOQFKRMKLNAKSPDVGVYKCOQLITDLIKOQIYER
IAHSNKSAGCIPIRYCKMKGLPYSCSMWDGAISLKRNPCITDEYSCNKOTPPK
DCVLVKRPRLVALKKOPSTYGSHGLELDHYOLNIMLAHMSKNCSCIADMGSI
KXTIOTISPLANTYFERHOIYGPRIYLVLSITLSMRQELQWASOMNAVYICDINSI
NMTRHEMMHQTKRKENTILLTYETILLDKAPLGALMAPIQVDPAHRAIKNDISLI
YKTLIDPKSNRIILITGSTPLONSIEISLHLHPKFSSEMPDFEEKGRVCYA
SLAKELEPFLLRVKRVKEKSLPAKVBDILRMKSKLOKYIKILTRYKLSKGSY
SGTSGLTIMMEYLRKCCNRCILKRPDNNEFYTKQDALHLKSSGKLILDKLILIR
REGKNRYLTISQWRMLDIJAELTKRFQFPQRLDSSIIGELRKQALDHFNBSGEDP
CFLLSTRAGGLGINLASDTVVYTFSDMNPNQDAQAARHTQOKQOVNIIRLVTKG
SVEBDILERAKKRWIDLHVLQIMDTGCTGLVTGASPSSTPRNKBELSAILKFGAB
ELEKEPEGSEODPEQIMIDEILKRAETHENPILSVDELISQFYANFSWMDDDI
ELEPERNSKNMBEILPEQORRLLEEEROKELBEITYLPNRNRKAQISFNSBERRS
RSRRYSGSDSDSISERRRPKRGKGRPRIPRENIKQSDEAIRRTISKYKFGCLPR
DATAARAELVMSKETDRLRIGELVINCGCYALKDSSSGTERAGARGKVGKPTGISG
```

VOVNAKLVIAHDELLIPLHKSIPSDPEERKOYITPCHTKAAHFIDIDWGEDSDNLLIG  
IYECYGSGMEMIKMPDLSTLTKILPDDPKQQAOLQTRADYIKLSRDLARREA  
ORLCGSGSMRKRTAKSKAMKSIKVEIKSDSPLESEDDDDKLNDSKPESEK  
DRSKSVSDADAVHTASGEVPVIAEESLEDOKEFSTCKEMRPVKALROLDBEK  
CLSEBOLEHTROCIKITGDHTTECLEKEISNEBOIKORKNIMITYSKRTEDARKLH  
KLTKAIAIKROESQONSDONSNVATTHVIRNPMERLENTNHDSSRDSYSDRHLS  
QYHDHNRHODSYKKSDSRKPYSFSGNDHEWHYRQDSRYSDREKHKRLD  
HRSRHRSLBEGGLDRCHSDHRSRSHSDHRSSEHTHHKSSRDYRRLSDWOLD  
HRAASSGRSPLDORSPYGSRSPFESHAHRSTPEHTWSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 187.00 Length: 41  
Ratio: 4.561 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.805

alignment\_block:

US-08-973-363-7 x MUSCHDIX ..

Align seg 1/1 to: MUSCHDIX from: 1 to: 5349

1 lileupProaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4026 ATTCTTCAGATGATCCTGATAAAAACCAACAAACAGTACAGAC 4075  
17 RATGALASPTYRLEULEUASNLYSASPLEUALAARGLYSG 34  
|||||  
4076 CCGTGCAGACTACCTCATCAACTACTTAGCAGAGATCTGCACAAAGAG 4125  
34 lualagInArgLeuAlaGlyAla 41  
|||||  
4126 AGGCTGACAGACTTGTGTGCGC 4148

seq\_name: gb\_pat:A58686

seq\_documentation\_block:

LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 153)

AUTHORS Griffiths, R. and Tiwari, B.

TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS

JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;

COMMENT ISIS INNOVATION (GB)

FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers

source 1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

alignment\_scores:

Quality: 182.00 Length: 51  
Ratio: 4.439 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 76.471

alignment\_block:

US-08-973-363-7 x A58686 ..

Align seg 1/1 to: A58686 from: 1 to: 153

1 lileupProaspProasp..... 7  
|||||

1 ATTTTACCTGATGCCAGATTAAGAAACCAACAGGCAAGCTTGACAGAC 50  
8 .LysLysProGlnAlaLysGlnLeuGlnThArgAlaAspTyrLeuLeu 24  
|||||  
51 CAAGAAACCAACAGGCAAGCTTGACAGACCGGTGCGAGATTACTCTTA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTAAGTAATTAAGACCTTGCAAGAAAGAGAGCTGACTGCT 150  
41 Ala 41  
|||  
151 GCG 153

seq\_name: gb\_pat:A58683

seq\_documentation\_block:

LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246

KEYWORDS

SOURCE unidentified.

ORGANISM unidentified.

REFERENCE 1 (bases 1 to 153)

AUTHORS Griffiths, R. and Tiwari, B.

TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS

JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;

COMMENT ISIS INNOVATION (GB)

FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers

source 1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 172.00 Length: 51  
Ratio: 4.195 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 70.588

alignment\_block:

US-08-973-363-7 x A58683 ..

Align seg 1/1 to: A58683 from: 1 to: 153

1 lileupProaspProasp..... 7  
|||||  
1 ATTTTTCAGATGATCCTGATAAAAACCAACCAACAGTACAGAC 50  
8 .LysLysProGlnAlaLysGlnLeuGlnThArgAlaAspTyrLeuLeu 24  
|||||  
51 CAAGAAACCAACAGGCAAGCTTGACAGACCGGTGCGAGACTACTCATCA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AACTACTTAGCAGAGATCTGCACAAAGAGAGGCTGACAGACTTGTGCT 150  
41 Ala 41  
|||  
151 GCG 153

seq\_name: gb\_pr:AF006513

seq\_documentation\_block:

LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997  
DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428



```

seq_name: gb_pr:AC012624
26832 AAGCT.....CTTCTGGTGCG 26816
34 lu1aglnargleua1aglala 41
||||| |||||
17 rrrgaalaaPTyrLeuIlleLysLeuLeuasnLysAspAlaAArgLysG 34
||||| ||||| ..... |||||
26882 CCGTCGACACTACCCTCATCAAAATTACTTAGTAGAGATCTTGCAAAAAAG 26833
34 lu1aglnargleua1aglala 41
||||| |||||
17 rrrgaalaaPTyrLeuIlleLysLeuLeuasnLysAspAlaAArgLysG 34
||||| ||||| ..... |||||
118297 CCGTCGACACTACCCTCATCAAAATTACTTAGTAGAGATCTTGCAAAAAAG 118346
34 lu1aglnargleua1aglala 41
||||| |||||
118347 AAGCT.....CTTCTGGTGCG 118363
seq_name: gb_htg:AC021449
seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
drive, Walnut Creek, CA 94598, USA
On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source location/Qualifiers
1..134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN
alignment_scores:
Quality: 170.00 Length: 41
Ratio: 4.359 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 85.366
alignment_block:
US-08-973-363-7 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365
1 ileleuproaspaspproaspyslylsgproglinalalyglleuglnrph 17
||||| ..... |||||
118247 ATTCTTCGCAGANGATCCGATAAAAACAACAACGAACAGTTGCAGAC 118296
17 rrrgaalaaPTyrLeuIlleLysLeuLeuasnLysAspAlaAArgLysG 34
||||| ||||| ..... |||||
118297 CCGTCGACACTACCCTCATCAAAATTACTTAGTAGAGATCTTGCAAAAAAG 118346
34 lu1aglnargleua1aglala 41
||||| |||||
118347 AAGCT.....CTTCTGGTGCG 118363

```

```

seq_documentation_block:
LOCUS      AC021449                143079 bp    DNA    linear    HTG-10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION  AC021449
VERSION    AC021449.3  GI:10047806
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 143079)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens, clone RP11-58M12
JOURNAL    Unpublished
AUTHORS     2 (bases 1 to 143079)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckery,R., Beda,F.,
            Bondurovsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepeel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Dearlana,P., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagsos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karalas,A., Klein,J.,
            Lechner,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
            MacDonald,P., Margulis,N., McEwan,P., McGurt,A., McKernan,K.,
            McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Pierre,N., Pisanl,C., Pollard,V., Raymond,C., Riley,R., Rotman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tzirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and zody,M.
COMMENT    Direct Submission
            Submitted (16-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 330 Charles Street, Cambridge, MA 02141, USA
            On Sep 10, 2000 this sequence version replaced gi:7407963.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L5154
            Center clone name: 58_M_12
            ----- Summary Statistics
            Sequencing vector: M13; M7815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap, version 0.960731
            Consensus quality: 134743 bases at least Q40
            Consensus quality: 139227 bases at least Q30
            Consensus quality: 140814 bases at least Q20
            Insert size: 14400; agarose-fp
            Insert size: 142179; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 38820: contig of 38820 bp in length
            * 38821 38920: gap of 100 bp
            * 38921 40411: contig of 1491 bp in length
            * 40412 40511: gap of 100 bp

```





Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366  
alignment\_block:  
US-08-973-363-7 x AC008531 ..  
Align seg 1/1 to: AC008531 from: 1 to: 145659  
1 11leuPrcAspAspProAspLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
46770 ATTCTTCACATATGATCCGATTAACCAACAGCAAAACAGTTGCAGAC 46819  
17 rAaGAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
46820 CCGTGAGACTACTCATCAATTAAGTAGAGAGATCTGCACAAAGAG 46869  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
46870 AACCT.....CTTCTGTGTCG 46886  
seq\_name: gb\_hcg:AC091946  
seq\_documentation\_block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRES \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCT-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-IP estimation  
Quality coverage: 5.38 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3824 3923: gap of unknown length

\* 3924 5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12460: contig of 2100 bp in length  
\* 12461 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14712 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18068: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20534: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 23615: gap of unknown length  
\* 23616 27563: contig of 3948 bp in length  
\* 27564 30987: gap of unknown length  
\* 30988 31087: contig of 3324 bp in length  
\* 31088 36065: gap of unknown length  
\* 36066 36165: contig of 4978 bp in length  
\* 36166 40978: gap of unknown length  
\* 40979 41078: contig of 4813 bp in length  
\* 41079 45663: contig of 4585 bp in length  
\* 45664 45763: gap of unknown length  
\* 45764 51745: contig of 5982 bp in length  
\* 51746 51845: gap of unknown length  
\* 51846 57359: contig of 5514 bp in length  
\* 57360 57459: gap of unknown length  
\* 57460 67981: contig of 10422 bp in length  
\* 67982 74132: gap of unknown length  
\* 74133 74232: contig of 6151 bp in length  
\* 74233 79795: gap of unknown length  
\* 79796 79895: contig of 5563 bp in length  
\* 79896 87511: gap of unknown length  
\* 87512 87611: gap of unknown length  
\* 87612 92791: contig of 5180 bp in length  
\* 92792 92891: gap of unknown length  
\* 92892 102794: contig of 9903 bp in length  
\* 102795 102894: gap of unknown length  
\* 102895 110866: contig of 7972 bp in length  
\* 110867 110966: gap of unknown length  
\* 110967 117571: contig of 6665 bp in length  
\* 117572 117671: gap of unknown length  
\* 117672 123738: contig of 6067 bp in length  
\* 123739 123838: gap of unknown length  
\* 123839 130583: contig of 6745 bp in length  
\* 130584 130684: gap of unknown length  
\* 130685 141544: contig of 10861 bp in length  
\* 141545 141644: gap of unknown length  
\* 141645 169109: contig of 27465 bp in length  
\* 169110 169209: gap of unknown length  
\* 169210 193446: contig of 24237 bp in length.  
Location/Qualifiers  
1..193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_lib="RPCT human BAC library 11"  
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN  
alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1







34 luA1aGlnArgLeuAlaGlyAla 41  
 ||||||||||||||||||  
 1238 AAGCAGACGAGACTTGCTGTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

AC AA142751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX Gallus sp.

FT Key Location/Qualifiers

FT CDS 228..5390

FT /\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 1; Fig 5; 76pp; English.

XX The chicken CHD-W gene (AA142754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AA142751) is suggested to initiate

CC female development in birds. The sequence of CHD-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AA142755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC gene (see also AA142756-57). It is located on an autosome or Z

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-tertile

CC bird and can be used for sex determination of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 205.00 Length: 41

Ratio: 5.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lIleuProAspProAspLysLysProGlnAlaLysGlnLeuIntH 17

4080 AATTTCCTCATGTCACAGACAAAGAAACCCAGCAAGCAGCTACAGAC 4129

17 rArgAlaAspTyrIleuIleuLeuLeuAsnLysAspLeuAlaArgLysG 34

||||||||||||||||||||||||||||||||||||||||  
 4130 CCGTGCAGACTACCTCATTAATTAAGTAAGACCTTGCAAGAAGC 4179

34 luA1aGlnArgLeuAlaGlyAla 41

||||||||||||||||||||  
 4180 AAGCACAAGGCTTGCTGTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq\_documentation\_block:

ID AA142757 standard; DNA; 153 BP.

AC AA142757;

DT 12-MAR-1997 (first entry)

DE Chick CHD-1A gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

FT Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),

CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)

CC and the great tit CHD-W gene (see also AA142759). Translated amino

CC acid sequences of this region are provided in AA08146-49. The

CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes

CC determine sex in birds and can be used to identify the sex of an

CC embryo, foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 190.00 Length: 51

Ratio: 4.634 Gaps: 1

Percent Similarity: 80.392 Percent Identity: 80.392

alignment\_block:

US-08-973-363-7 x AA142757 ..

Align seg 1/1 to: AA142757 from: 1 to: 153

1 lIleuProAspProAsp..... 7

```

|||||
1 ATTTACTGATGATGATCCAGACAGAAACCCAGCAAGCAGCTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleI 24
|||||
51 CAAGAAACCCAGCAAGCAGCTACAGACCCCTGCGACATTCCTCATTA 100
24 yslleuLeuAsnLysAspLeuAlaArgLysGlnAlaArgLeuAlaGly 40
|||||
101 AATTACTGATAATTAAGACCTTGCGAAGAAAGACACACAGACTTGCTGGT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
AC AA142758;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chick CHD-W gene fragment.
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81 /tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI: 1997-043127/04.
XX
DR P-PSDB; AAW08148.
XX
PA Avian chromodomain-helicase-DNA binding genes determine sex in
XX PT birds - used for sex determ. and to control sex of progeny
XX PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC and the great tit CHD-W gene (AA142759). Translated amino acid
CC sequences of this region are provided in AAW08146-49. The CHD-1A
CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX
SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Alignment_scores:
Quality: 190.00 Length: 51
Ratio: 4.634 Gaps: 1

```

```

Percent Similarity: 80.392 Percent Identity: 80.392
Alignment_block:
US-08-973-363-7 x AA142758
Align seg 1/1 to: AA142758 from: 1 to: 153

1 IleLeuProAspAspProAsp.....7
1 ATTTACTGATGATGATCCAGTAAAGAAACCCAGCTACAGCTTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleI 24
|||||
51 CAAGAAACCCAGCAAGCAGCTACAGACCCCTGCGACATTCCTCATTA 100
24 yslleuLeuAsnLysAspLeuAlaArgLysGlnAlaArgLeuAlaGly 40
|||||
101 AATTACTGATAATTAAGACCTTGCGAAGAAAGACACACAGACTTGCTGGT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759
seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
AC AA142759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81 /tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI: 1997-043127/04.
XX
DR P-PSDB; AAW08149.
XX
PA Avian chromodomain-helicase-DNA binding genes determine sex in
XX PT birds - used for sex determ. and to control sex of progeny
XX PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC and the great tit CHD-W gene (AA142759). Translated amino acid
CC sequences of this region are provided in AAW08146-49. The CHD-1A
CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,

```



CC foetus etc. and to manipulate the sex of progeny.  
 XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment\_scores:  
 Quality: 182.00 Length: 51  
 Ratio: 4.439 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 76.471

alignment\_block:  
 US-08-973-363-7 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

```

1 1leuPProAspProasp..... 7
1 ATTTTACCTGATGATCCGATTAAGAAACACAGCAAGCAAGTTCAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
51 CAGAAACACAGCAAGCAAGTTCAGACCCGTCGACATTACTCATTA 100
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AATTAAGTAATTAAGACCTTGCAAGAAAGAAAGTCAAGACTTACTGT 150
41 Ala 41
151 GCA 153

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756

seq\_documentation\_block:

ID AAT42756 standard; DNA; 153 BP.

XX AAT42756;

DT 12-MAR-1997 (first entry)

DE Mouse CHD-1 gene (bases 3855-977).

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1; CHD-W; W chromosome; ss.

OS Mus sp.

EH Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 and are ignored in the translated amino  
 acid sequence given in Fig 3"

XX WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAM08146.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
 CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
 CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
 CC Translated amino acid sequences of this region are provided in  
 CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
 CC AAT42754-55) genes determine sex in birds and can be used to identify  
 CC the sex of an embryo, fetus etc. and to manipulate the sex of  
 CC progeny.

XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

alignment\_scores:  
 Quality: 172.00 Length: 51  
 Ratio: 4.195 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 70.588

alignment\_block:  
 US-08-973-363-7 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

```

1 1leuPProAspProasp..... 7
1 ATTTTACCTGATGATCCGATTAAGAAACACAGCAAGCAAGTTCAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
51 CAAAAAACACACAGCAAGCAAGTTCAGACCCGTCGACATTACTCATTA 100
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AACTACTTACGACGATCTTGCAAAAAAGAGAGCTCAGACACTTGTGT 150
41 Ala 41
151 GCG 153

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; cDNA; 1311 BP.

XX AAV59280;

DT 14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

XX Homo sapiens.

OS Synthetic.

EH Key Location/Qualifiers

FT CDS 1..1311

FT /tag= a

FT /product= "A-TRF"

PD WO9836066-A1.

PF 13-FEB-1998; 98WO-US02765.

PR 04-FEB-1998; 98US-0018628.

XX 13-FEB-1997; 97US-0800264.  
 XX (UVRQ ) UNITV ROCKEFELLER.  
 XX

XX	PI	Blanchi A, De Lange T, Van Steensel B;	PR	17-MAR-2000;	200005-019007	PR
XX	DR	WPI; 1998-480769/41.	PR	18-APR-2000;	200005-0198123	PR
XX	P-PSDB;	AAW59280.	PR	19-MAY-2000;	200005-0205515	PR
PT	XX	Nucleic acid encoding altered telomere repeat binding protein and related vectors - transformants, hetero-dimers and antihodis, used to inhibit shortening of telomerases caused by ageing or disease,	PR	07-JUN-2000;	200005-0209467	PR
PT	XX	also used to extend life of cells in culture	PR	28-JUN-2000;	200005-0214886	PR
PS	XX	Claim 14; Page 110-111; 163pp; English.	PR	07-JUL-2000;	200005-0215135	PR
CC	XX	The altered vertebrate telomere repeat binding protein (A-TRF) has a	PR	07-JUL-2000;	200005-0216680	PR
CC	XX	telomere repeat binding factor (TRF) dimerisation domain, and forms a	PR	11-JUL-2000;	200005-0217495	PR
CC	XX	hetero-dimer with TRF, preventing it from binding to the specified repeat	PR	26-JUL-2000;	200005-0218920	PR
CC	XX	sequence. A-TRF, optionally expressed by gene therapy, is used to	PR	26-JUL-2000;	200005-0220563	PR
CC	XX	inhibit shortening of telomeres associated with ageing (for cosmetic	PR	14-AUG-2000;	200005-0220964	PR
CC	XX	purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,	PR	14-AUG-2000;	200005-0224518	PR
CC	XX	atrophy of the skin, age-related macular degeneration, atherosclerosis,	PR	14-AUG-2000;	200005-0224519	PR
CC	XX	cellular and viral (including human immune deficiency virus) infection.	PR	14-AUG-2000;	200005-0225213	PR
CC	XX	Cells expressing A-TRF also have an increased life span in vitro, e.g.	PR	14-AUG-2000;	200005-0225214	PR
CC	XX	for expression of recombinant proteins or where intended for subsequent	PR	14-AUG-2000;	200005-0225267	PR
CC	XX	transplant or for testing, eliminating the need for transformation.	PR	14-AUG-2000;	200005-0225282	PR
SQ	XX	Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;	PR	14-AUG-2000;	200005-0225270	PR
			PR	14-AUG-2000;	200005-0225447	PR
			PR	14-AUG-2000;	200005-0225757	PR
			PR	14-AUG-2000;	200005-0225758	PR
			PR	18-AUG-2000;	200005-0226271	PR
			PR	22-AUG-2000;	200005-0226681	PR
			PR	22-AUG-2000;	200005-0226682	PR
			PR	23-AUG-2000;	200005-0227182	PR
			PR	30-AUG-2000;	200005-0228924	PR
			PR	01-SEP-2000;	200005-0229287	PR
			PR	01-SEP-2000;	200005-0229343	PR
			PR	01-SEP-2000;	200005-0229344	PR
			PR	01-SEP-2000;	200005-0229345	PR
			PR	05-SEP-2000;	200005-0229509	PR
			PR	06-SEP-2000;	200005-0230437	PR
			PR	08-SEP-2000;	200005-0230438	PR
			PR	08-SEP-2000;	200005-0231242	PR
			PR	08-SEP-2000;	200005-0231243	PR
			PR	08-SEP-2000;	200005-0231244	PR
			PR	08-SEP-2000;	200005-0231414	PR
			PR	08-SEP-2000;	200005-0232080	PR
			PR	08-SEP-2000;	200005-0232081	PR
			PR	12-SEP-2000;	200005-0231968	PR
			PR	14-SEP-2000;	200005-0232397	PR
			PR	14-SEP-2000;	200005-0232398	PR
			PR	14-SEP-2000;	200005-0232399	PR
			PR	14-SEP-2000;	200005-0232401	PR
			PR	14-SEP-2000;	200005-0232401	PR
			PR	14-SEP-2000;	200005-0233063	PR
			PR	14-SEP-2000;	200005-0233064	PR
			PR	14-SEP-2000;	200005-0233065	PR
			PR	21-SEP-2000;	200005-0234223	PR
			PR	21-SEP-2000;	200005-0234224	PR
			PR	21-SEP-2000;	200005-0234224	PR
			PR	25-SEP-2000;	200005-0234997	PR
			PR	25-SEP-2000;	200005-0234998	PR
			PR	26-SEP-2000;	200005-0235484	PR
			PR	27-SEP-2000;	200005-0235834	PR
			PR	27-SEP-2000;	200005-0235835	PR
			PR	29-SEP-2000;	200005-0236327	PR
			PR	29-SEP-2000;	200005-0236368	PR
			PR	29-SEP-2000;	200005-0236369	PR
			PR	29-SEP-2000;	200005-0236369	PR
			PR	02-OCT		

PR 13-OCT-2000; 2000US-02393937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251719.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251899.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCT INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI: 2001-502630/55.  
DR P-PSDB; AAM93109.  
XX  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosis, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
XX  
PS Claim 1; SEQ ID NO 1198; 986bp; English.  
XX  
CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other:  
  
alignment\_scores:                   Length:       35  
                  Quality:   122.00  
                  Ratio:     4.067  
Percent Similarity: 85.714   Percent Identity: 71.429  
  
alignment\_block:  
US-08-973-363-7 x AAK88882 ..  
  
Align seq 1/1 to: AAK88882 from: 1 to: 421  
  
1   lileuProAspProAspLysLysProGlnAlaLysGlnLeuGlnH 17  
      |||||  
110   ATTCCTGCGCGTGGAGACAGATAAAGCCTCAGGGGAGAGCTACGAC 159  
      |||||  
17   rAtgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
      |||||  
160   CCGAGCGGATTACTTGTGTAAGCTGCTCAGAAAGCGTCTGGAGAGAGAGG 209  
      |||||  
34   luAla 35  
      |||  
210   GGGCT 214  
  
seq\_name: /stsl/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA157603  
seq\_documentation\_block:  
ID   AA157603 standard; cDNA; 421 BP.  
XX  
AC   AA157603;  
XX  
DT   19-OCT-2001 (first entry)  
XX  
DE   Human colorectal cancer antigen cDNA SEQ ID NO: 67.  
XX  
KW   Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
XX  
OS   Homo sapiens.  
XX  
PN   WO200155350-A1.  
XX  
PD   02-AUG-2001.  
XX  
PF   17-JAN-2001; 2001WO-US01350.  
XX  
XX  
31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.



Percent Similarity: 85.714 Percent Identity: 71.429  
 alignment\_block:  
 US-08-973-363-7 x AAI57603 ..

Align seg 1/1 to: AAI57603 from: 1 to: 421

```

1 lleuProaspaspProaspLysProGlnAlaLysGlnLeuGlnH 17
  |||||  ::  |||||  ::  |||||  ::  |||||  ::
110 ATTCTGCGGTTGAGACAGATATAAGCTCAGGAGAGACGCTACAGAC 159
17 rArgAlaAspTyrIleuIleLysIleuAsnLysAspLeuAlaArgLysG 34
  |||||  ::  |||||  ::  |||||  ::  |||||  ::
160 CCGAGCGGATTACTTGTGAAGCTGCTCAGAAAGGCTGAGAGAAGG 209
34 luAla 35
  |||
210 GGCGT 214
  
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:  
 ID ABL06443 standard; cDNA: 6240 BP.

```

XX AC ABL06443;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB62340.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX CC sequences (ABLI01840-ABLI6175) and the encoded proteins
XX CC (ABBS7737-ABBS72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;
  
```

alignment\_scores:

Quality: 88.00 Length: 38  
 Ratio: 2.839 Gaps: 2  
 Percent Similarity: 81.579 Percent Identity: 55.263

alignment\_block:  
 US-08-973-363-7 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```

1 lleuProaspaspProaspLysProGlnAlaLysGlnLeuGlnH 17
  |||||  ::  |||||  ::  |||||  ::  |||||  ::
4184 ATTTTGCTTAACGAT...ACCGCAAGCCCCAGGCAAGAGCTGACGAC 4230
17 rArgAlaAspTyrIleuIleLysIleuAsnLysAspLeu...AlaArgL 33
  |||||  ::  |||||  ::  |||||  ::  |||||  ::
4231 GCGTCCGAGTACTGCTCAAGATCATCAAGAGACGTGAGCTGACCA 4280
33 ysgLuAlaGlnArg 37
  ||  ::  |||||
4281 AGGAGAGCAACACG 4294
  
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:  
 ID ABL06442 standard; cDNA: 9933 BP.

```

XX AC ABL06442;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB62339.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX CC sequences (ABLI01840-ABLI6175) and the encoded proteins
XX CC (ABBS7737-ABBS72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;
  
```









**THIS PAGE BLANK (USPTO)**



seq\_documentation\_block: 619 bp mRNA linear EST 18-Oct-2001

LOCUS B155356 Musculus full-length enriched, 16 days neonate thymus Mus

DEFINITION musculus cDNA A130024f16.3' similar to I10410 Mouse

DNM-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION B155356

VERSION B155356.2 GI:16268254

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 619)

AUTHORS Arakawa,T., Carinci,P., Fukuda,S., Furuno,M., Hanaagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Komoto,H., Kouda,M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okito,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Jun 29, 2000 this sequence version replaced gi:8811286.

Contact: Yoshinobu Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gscc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
 Nagai, K., Fujiyama, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multichannel sequencer. *Genome Res.*  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa  
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences *Mamm. Genome* 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Location/Qualifiers  
 1. 619  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A130024L16"  
 /clone\_lib="RIKEN full-length enriched, 16 days neonate  
 thymus"  
 /issue\_type="thymus"  
 /dev\_stage="16 days neonate"  
 /lab\_host="DH10B"  
 /note="Site\_1: Saito; Site\_2: BamHI. cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGAGATCCACAGACCTCTTTTTTTTTTTTTVVN 3'], cDNA was  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATTCGAGTTAAATTAATATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I. "

BASE COUNT	194	a	118	c	161	g	146	t
ORIGIN								
alignment_scores:								
Quality:	187.00							Length: 41
Ratio:	4.561							Gaps: 0
Percent Similarity:	100.000							Percent Identity: 87.805
alignment_block:								
US-08-973-363-7	x	BB155356	..					
Align seg 1/1	to:	BB155356	from:	1	to:	619		

```

1      1  lleleupProaspaspProaspLysLysProGlnAlaLysGlnLeuGln 17
489  ATTCTTCAGATGATCTGATATAAAACACACAGCAAAACACTTCACAGC 538
17  TATGAAlaasPTyrlLeuileLysLeuLeuAsnLysaspLeuAlaArgLysG 34
539  CGGTGCAGACTACCTCATCAACTACTTACGACAGAGACTTGCAGAAAGAG 588
34  TATAGlnArgLLeuAlaGlyAla 41
589  AGGCTCAGAGACTTGTGTGTGCG 611

seq_name: gb_est1:AL644594

seq_documentation_block:
LOCUS      AL644594                645 bp      mRNA      linear      EST 07-NOV-2001
DEFINITION      AL644594 XGC-egg silurana tropicalis cDNA clone L1E1d12 5', mRNA
sequence.
ACCESSION      AL644594
VERSION        AL644594.1  GI:16796719
KEYWORDS
SOURCE
ORGANISM      western clawed frog.
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE      1 (bases 1 to 645)
AUTHORS      Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE      Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL
COMMENT      Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: L1E1d12.plc
Sequencing primer: PLC
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..645
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="L1E1d12"

```

```

/clone_11b="XGC-egg"
/dev_stage="egg"
/lat_host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

BASE COUNT      222 a      125 c      156 g      141 t      1 others

ORIGIN

alignment_scores:
      Quality: 187.00      Length: 39
      Ratio: 4.921      Gaps: 0
      Percent Similarity: 97.436      Percent Identity: 94.872

alignment_block:
US-08-973-363-7 x AL644594 ..

Align seq 1/1 to: AL644594 from: 1 to: 645

1 11leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
456 ATTTTACCAGATGATCCAGATTAAAGAACCCCAAGCAAGCATGATACAC 505
|||||
17 TATGAlAspTyrLeuLeuLysLeuAsnLysAspLeuAlaArgLVSG 34
|||||
506 CAGAGCTGATACCTCATTTAAACTTCTCAATTAAGATCTGTTAGGAAG 555
|||||
34 lvaLglnArgLeuAla 39
|||||
556 AAGCACAAAGACTTCT 572

seq_name: gb_estl:BB461065

seq documentatn block:
LOCUS      BB461065              660 bp      mRNA      linear      EST_25-OCT-2001
DEFINITION      BB461065 RIKEN full-length enriched, 12 days embryio spinal ganglion
               Mus musculus cDNA clone D130070513.3 similar to l10410 Mouse
               DNA binding protein (Cbp-1) mRNA, mRNA sequence.
ACCESSION      BB461065.2 GI:16426612
VERSION        BB461065
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus
               Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 660)
REFERENCE      1
AUTHORS       Areakawa,T., Carrincci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A,A
               , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,K.,
               'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
               'Ozaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasak,K.,
               'D., Shibata,K., Shinagawa,A., Shitaki,T., Sogabe,Y., Suzuki,H.,
               'Tegami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,
               'Muramatsu,M. and Hayashizaki,Y.
               RIKEN Mouse ESTs (Arkawa,T., et al. 2001)
               Unpublished (2001)
TITLE         Not published (2001)
JOURNAL
COMMENT       Contact: Yoshitake Hayashizaki
               Laboratory for Genome Exploration Research Group, RIKEN Genomic
               Sciences Center(GSC), Yokohama Institute
               The Institute of Physical and Chemical Research (RIKEN)
               1-7-2 Shubito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               Tel: 81-45-503-9222
               Fax: 81-45-503-9216
               Email: genome-res@sc.riken.go.jp,
               urh:http://genome.gsc.riken.go.jp/
               'K., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
               Normalization and subtraction of cap-trapped-selected cDNAs to
               prepare full-length cDNA libraries for rapid discovery of new
               genes. Genome Res. 10 (10), 1617-1630 (2000)
               wagi,K., Fujiwake,S., Inoue,K., Yogawa,Y., Izawa,M., Ohara,E.,

```

```

MATCHIKI,M., YONEDA,Y., ISHIKAWA,T., OZAWA,K., TANAKA,T., MATSUURA
,S., KAWAI,J., OKAZAKI,Y., MURAMATSU,M., INOUE,Y., KITRA,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carmichael,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kono,S., Shingara,A., Saito,T., Kiyosawa,H., Yamahara,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.ysc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
source
1. 660
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_1b="D13007B13"
/clone_1lb="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH108"
/note="Site.1: Salt; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGCGGCGCCACTCGAGCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using Tretalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAGACTTCGAGTTAAATTAATTAATTCCTCCGCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
phluescript KS(+) after bulk excision from lambda f1d I."
BASE COUNT
211 a 130 c 168 g 151 t
ORIGIN
Alignment_scores:
Quality: 187.00 Length: 41
Ratio: 4.561 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.805
Alignment block:
US-08-973-363-7 x BBA61065 ..
Align seg 1/1 to: BBA61065 from: 1 to: 660
1 lleleuproaspaspProksplyslsyrproGlnalysGlnenclnlt 17
|||||
481 ATTCTTCCAGATGATCTCTGATATTAACCAACGAAACGATTACAGC 530
|||||
17 TATGAlAspTytleuilelyslleuAsnlyslAspleuAlaGlySG 34
|||||
531 CGGTGAGACGACCTCATCAACTTCTTACGACGAGATCTGCATAAAGG 580
|||||
34 lnaAclnaryleuAlaGlyAla 41
|||||
581 AGGCTGACAGACTTGTGGTGGC 603

```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001

DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527

Yjg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA sequence.

ACCESSION BB834922

VERSION BB834922.1 GI:17013165

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 446)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscl.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

FEATURES

source Location/Qualifiers

1..446

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="G930033J21"

/clone\_1lb="RIKEN full-length enriched, mammary gland RCB-0527 Yjg-MC(B) cDNA"

/tissue\_type="mammary gland"

/cell\_line="RCB-0527 Yjg-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores: Length: 41

Quality: 182.00

Ratio: 4.439

Percent Similarity: 100.000 Percent Identity: 82.927

alignment\_block:

us-08-973-363-7 x BB834922

Align seg 1/1 to: BB834922 from: 1 to: 446

1 l1leuPProaspPProaspLysPProGlnAlaLysGlnLeuGlnIth 17

|||||

309 ATTCTCCAGATGATCCTGATTAACCAACCAACAGGTTACAGAC 358

17 rargAlaAspTyrIleuLeuLeuLeuAsnLysAspLeuAlaArgLysG 34

|||||

359 CCGTCAGACTACCTCCTCAACTACTAGCAGATCTTGCCAAAGAG 408

34 luAlaGlnArgLeuAlaGlyAla 41

|||||

409 AGGCTCAGACACTTGTGCTTCG 431

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001

DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527

Yjg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA sequence.

ACCESSION BB830730

VERSION BB830730.1 GI:17008973

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 438)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscl.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

```

FEATURES
  source
    Location/Qualifiers
      1..438
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_id="G930013K04"
        /clone_1lb="RIKEN full-length enriched, mammary gland
        RCB-0527 Jy9-MC(B) cDNA"
        /tissue_type="mammary gland"
        /cell_line="RCB-0527 Jy9-MC(B)"
        /origin
          138 a 96 c 108 g 96 t

  alignment_scores:
    Quality: 172.00 Length: 42
    Ratio: 4.195 Gaps: 1
    Percent Similarity: 97.619 Percent Identity: 83.333

  alignment_block:
    US-08-973-363-7 x BB830730 ..
    Align seg 1/1 to: BB830730 from: 1 to: 438

  seq_documentation_block:
    LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001
    DEFINITION DKFZP313J1040.r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
    DKFZP313J1040 5', mRNA sequence.
    ACCESSION AL601246
    VERSION AL601246.1 GI:15164752
    KEYWORDS EST.
    SOURCE
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      Bloeker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann
      S. (Bloeker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann
      S.)
      Unpublished (1999)
      Contact: Bloeker, H
      MRP
      Am Klopferstr. 18a D-82152 Martinsried, Germany
      This is the 5' sequence of the clone insert
      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
      Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
      sequenced by GBR (National Research Centre for Biotechnology Ltd.,
      Braunschweig/Germany) within the cDNA sequencing consortium of the
      German Genome Project.
      No 5' sequence available.
      This clone (DKFZP313J1040) is available at the RZPD in Berlin.
      Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059
      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
      Location/Qualifiers
        1..430
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_id="DKFZP313J1040"
          /clone_1lb="313 (synonym: hlcc2)"

FEATURES
  source
    /dev_stage="adult"
    /lab_host="DH10b"
    /note="Vector: pTriplex2; Site_1: SfilA; Site_2: SfilB;
    cDNA collection"
    BASE COUNT 163 a 81 c 86 g 100 t
    ORIGIN
      163 a 81 c 86 g 100 t

  alignment_scores:
    Quality: 170.00 Length: 41
    Ratio: 4.359 Gaps: 1
    Percent Similarity: 95.122 Percent Identity: 85.366

  alignment_block:
    US-08-973-363-7 x AL601246 ..
    Align seg 1/1 to: AL601246 from: 1 to: 430

  seq_documentation_block:
    LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000
    DEFINITION Wm55f11.x1 NCI-CGAP-U12 Homo sapiens cDNA IMAGE:2443725 3'
    similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
    PROTEIN 1; , mRNA sequence.
    ACCESSION A1890775
    VERSION A1890775.1 GI:5595939
    KEYWORDS EST.
    SOURCE
      ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
      Unpublished (1997)
      Contact: Robert Strausberg, Ph.D.
      Email: cgaps@email.nih.gov
      Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
      Emmert-Buck, M.D., Ph.D.
      CDNA Library Preparation: Life Technologies, Inc.
      DNA Sequencing by: Greg Lennon, Ph.D.
      Clone distribution: NCI-CGAP clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      www-bio.lnl.gov/dbfp/image/image.html
      Insert Length: 1924 Std Error: 0.00
      Seq primer: -400P from Gibco
      High quality sequence stop: 418.
      Location/Qualifiers
        1..547
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_id="IMAGE:2443725"
          /clone_1lb="NCI-CGAP-U12"
          /tissue_type="moderately-differentiated endometrial
          adenocarcinoma, 3 pooled tumors"
          /lab_host="DH10b"
          /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
          site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

```

Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 114 a 118 c 85 g 230 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-7 x A1890775/rev ..

Align seg 1/1 to reverse of: A1890775 from: 1 to: 547

seq\_name: gb\_est1:AUI25712

seq\_documentation\_block:  
LOCUS AUI25712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AUI25712 NT2RM4 Homo sapiens cDNA clone NT2M4002061 5', mRNA  
sequence.  
ACCESSION AUI25712  
VERSION AUI25712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
,Y., Sugano,S., Isogai,T.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source  
1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-7 x AUI25712 ..

Align seg 1/1 to: AUI25712 from: 1 to: 866

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:  
LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1028)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: L1AM9753 row: h column: 16  
High quality sequence stop: 488.  
Location/Qualifiers  
1. 1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Skin; Vector: PCMV-SPOrt6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 387 a 205 c 238 g 198 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:





Align seg 1/1 to: BF239967 from: 1 to: 821

```
6 ProasplyslysglnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22
|||||
3 CCCGATATAAAACCAACAGCAAGTTCGACAGCCCTGCAGACTTACT 52
|||||
22 uilLeuLeuLeuAsnLysAspLeuAlaArgLysGlnuLaglnArgLeuA 39
|||||
53 CATCAATTAATCTAGTAGAGATCTGCAGAAAAAGAACT.....CTTT 96
|||||
39 lagllyAla 41
|||||
97 CTGCTGCG 104
```

seq\_name: gb\_est1:AM996787

seq\_documentation\_block:

LOCUS AM996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM996787  
VERSION AM996787.1 GI:8257021

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230

200-102-d03&tl=2000-02-23&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

Location/Qualifiers

1..337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI: A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores: Quality: 133.50 Length: 36  
Ratio: 3.926 Gaps: 2  
Percent Similarity: 94.444 Percent Identity: 83.333

alignment\_block:

us-08-973-363-7 x AM996787/rev ..

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

```
1 lLeuProAspAspProAsplyslysglnAlaLysGlnLeuGlnThr 17
|||||
105 ATTCTTCAGATGATCCGATATAAAACCAACAGCAAGTTCGACAG 56
|||||
17 rArgAlaAspTyrLeuileLys.LeuLeuAsnLysAspLeuAlaArgLys 33
|||||
55 CCGTCAGAGACTATCTCATCACTT...AGTAGAGATCTTGCAGAAAAA 9
|||||
34 GluAla 35
|||||
8 GAACT 3
```

seq\_name: gb\_est1:AM997058

seq\_documentation\_block:

LOCUS AM997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM997058  
VERSION AM997058.1 GI:8257292

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150

400-152-c03&tl=2000-04-15&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 19

High quality sequence stop: 678.

Location/Qualifiers

1..686

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI: A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t

ORIGIN

alignment\_scores: Quality: 133.50 Length: 36  
Ratio: 3.926 Gaps: 2  
Percent Similarity: 94.444 Percent Identity: 83.333



---



source 1..6608 /organism="unidentified" /db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN  
alignment\_scores:  
Quality: 205.00 Length: 41  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608  
1 lleleuProaspProaspLysProglInAlalysGlnleuGlnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGCCTTGCAGAGAGG 4129  
17 fArgAlaSPtyrLeuIleuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAGG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAGGCTTGCTGTGTGCA 4202  
seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus chromo-helicase-DNA-binding on the z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cds  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 6872)  
REFERENCE  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE A CHD1 gene is z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE  
2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/translation="MNGHSDSESVNNSGESSRDDSGASGSGSSGSSGSS  
SOSGSDSESGESGSESSESDTSREKKQYAKPKADGSEFMKSPETILAVQSAV  
LKKOQOQKAASDSGSESDSDSDADSDSETKKKKHDDQMSGSGSVSGTGS  
DSEADGDKSSCESSESDYEPKRVKSKRPRIKPKSGKSTGQKKRQDLSSEEE

DDDEDYDKGRSRRQATVNVSYKAEETKTSDDLLEVCEDVPQTEDEFEETIEKFM  
SRIGRGATGASTTIIYAAVEDDGPVAGCEKSELEIYUJLKKMGSHIHMTWEET  
LKOQVKKGNKLDNTYKADQETKRWLKNASPDVYVYCOOELTPDLKHOYVIRI  
AHSNOKSAAGPDYCKMGOGLPESCSMDGALIAKKQARIDEVFSNOSKTPFKD  
CYLQKQRPYALAKQPSYIGHESELELDYOLNGLNLASHKCKGNCILADGELG  
KTQITSPNLVLEHOLGEPFLVPLSTLTSWQREIQVMAPOKNVAVYIGDITSRN  
MIRTHEMHPQTRKLFNILLTYEILLDKSEFLGLMNAFYGDEARLKNDSLY  
RTLIDFKSNHRLITGTPLONSLKELWSLHEIMPEKFSWDEPEEGKRGYAS  
LHKELEPFLRRVKQVERSLPAKVEQLRMMESALQOYKWIITRYKALKSGSK  
STSGFLIMMELKCCNHCYLKPPDDNPEFYKQALOHILRSSGLILDLKLRLR  
PQGNVLIFFSQWVRMLDILAEYLYKRPORFQDSDSTIGELRKAOLDHFNMGSDPC  
FLSTRAGGLGINTLASDTVYITDSDMPQNDLOQANARHIGQKQYNIIRLYTKGS  
VEEDILERRAKKMWLDHLVLOMDTGTGTVLHTGSTPSSSTPFNKEELAILKFAE  
LEKEPEGEEOEPQMDIDELIKRAETRENEPGLVYGBELLQFQVAFSNVDEIDIE  
LEPERNSRMWEEIIPESORRIIEEEROKELEIYMLPRMRCAQISFNSGSRSR  
SRRYSQSDSDSTTERPRPKRGRRPTIREPKGSDAIRRFITSYKFGPGLERLD  
AVARDAELVDKSEPTQRLRGLVHNCITKALKDSSGOBRGRLGKGPFRISGV  
QVNAKLVISHEELAPLHKSLSPDEEKKRYTPCHTAAHEDITMGEDDSNLVGI  
YEGYGSWEMIIMKDPDLSLTLQILPDDPKKQQAOLQTRADYLLIKLNLKDLAREAQ  
FLAGAGSNKRRKTRNNKMKASKIREEIKSDSSPQSPSEKSEDEEDDNKDEIVSK  
HLKKIKTEKEENKEEPDIDIGIKFAEERETKEKREKRELKREKEDKELKED  
NKEKRENNKESSTOKETKYEKEKVMKSEKKEKKTPLDTPVHTATSPPISE  
ESELHOKTFSYCKERMRYKALKQIDRPEGLSEPRQLEHTRCIIKIGHTIECL  
KEYTPEQTKQKRNLMLEFSKFTFEDKALHLKYKHAIRKQESQAHNDQVNSVN  
THVIRNPVERLKETTNHDDSSRDYSDDRLLSQYHDDHKKDHQDAVKKSDSRPY  
SAFSNGKDHDMWDHYKODSRYSYSDSKHRLDHRSDRHHNLEGMKIDSRGSHDRSH  
SDRLHSDHRSSTSEYSHKSSRDYRHSQWQMDHRASSGSPRSPLDQSPSYGSRPLG  
HRSPEHSDHKSSTPEHWTSSKKT"  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
alignment\_scores:  
Quality: 205.00 Length: 41  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872  
1 lleleuProaspProaspLysProglInAlalysGlnleuGlnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGCCTTGCAGAGAGG 4129  
17 fArgAlaSPtyrLeuIleuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAGG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAGGCTTGCTGTGTGCA 4202  
seq\_name: gb\_pat:A58684  
seq\_documentation\_block:  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
KEYWORDS A58684.1 GI:371247  
SOURCE  
ORGANISM  
unidentified.  
unidentified.  
REFERENCE  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;

COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 source 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 58 a 40 c 31 g 24 t  
 ORIGIN

Alignment\_scores:  
 Quality: 190.00 Length: 51  
 Ratio: 4.634 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 80.392

Alignment\_block:  
 US-08-973-363 x A58684 ..

Align seq 1/1 to: A58684 from: 1 to: 153

1 HleuProaspasProasp..... 7  
 1 ATTTTACTGATGATCCAGACCAAGAACCCAGGCAAGACGATACAGAC 50  
 1 ATTTTACTGATGATCCAGACCAAGAACCCAGGCAAGACGATACAGAC 50  
 8 LysLysProGlnAlaLysGlnLenglnThrArgAlaAspTyrLeuLeu 24  
 1 ATTTTACTGATGATCCAGACCAAGAACCCAGGCAAGACGATACAGAC 50  
 51 CAGAAACCCAGGCAAGACGATACAGACCCCTGACGATACCTATTA 100  
 24 yslleuAenlysaspleuAlaArgylsGluAlaArgLeuAlagly 40  
 101 AATTACTGATTAAGACCTTGACAGAAAGAGACCAAGGCTTGCTGCT 150  
 41 Ala 41  
 151 GCA 153

seq\_name: gb\_pat:A58685

seq\_documentation\_block:  
 LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 4 from Patent WO9639505.  
 ACCESSION A58685  
 VERSION A58685.1 GI:3714248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 Griffiths, R. and Tiwari, B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
 ISIS INNOVATION (GB)  
 COMMENT Other publication AU 5906996 961224.  
 FEATURES  
 source 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t  
 ORIGIN

Alignment\_scores:  
 Quality: 190.00 Length: 51  
 Ratio: 4.634 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 80.392

Alignment\_block:  
 US-08-973-363 x A58685 ..

Align seq 1/1 to: A58685 from: 1 to: 153

1 HleuProaspasProasp..... 7

VOVNAKLVIAHEDELLIPLHKSTPSPDEERKQYITPCHTKAHFIDWGEDSDNLLIG  
IYEVYGSWEMITKMDPDLSTJHKILPDDPKRPAKOLQTRADYILIKLSRLAKREA  
ORLCGAGSGMKTTRAKRSKAMKSIKKEIKSDSPLESEKSDDDDKLNDKLRREA  
DESKSVSDAPVHTASGEVPIAESEELPOKFSICKEMRKYKALAKOLDPEK  
GLSEEROLEHTPFOCLIKIGDHTTECLKEYSNEOLKOWRKNMIVYSKPTPEADAKLH  
KLXKAIAIKKROSDQONSNDONSATVTHYIRNDMERLKNTHHDSRSDSYSDRHLIS  
QTHDHKDHQSDYKSKSDSRKPYSSFSNGDKHEDWBYRODSRYSDREKHRLKD  
HRSRHRPSLEGLDRCHSDHRSYSDHMSHSDHSSSEHTHKSRSRYRYLSDWOLD  
HRAASSGPRSPIDORSPYGSRSPFESHAHRSTPEHTWSSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

## alignment\_scores:

Quality: 187.00 Length: 41  
Ratio: 4.561 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.805

## alignment\_block:

US-08-973-363-8 x MUSCHDX ..

Align seg 1/1 to: MUSCHDX from: 1 to: 5349

1 l1leuProaspProaspPlysPlysProglAlaLysGlnLeuGlnTh 17  
|||||  
4026 ATTCTTCACAGATGATCCTGATATAAACACAAACAAAGCTTACAGAC 4075  
|||||  
17 l1leuProaspPlysPlysProglAlaLysGlnLeuGlnTh 34  
|||||  
4076 CCGTGCAGACTACCTCATCAACTACTTAGCAGATCTTGCAAAAGAG 4125  
|||||  
34 l1leuProaspPlysPlysProglAlaLysGlnLeuGlnTh 41  
|||||  
4126 AGGCTCAGAGACTTGTGTGCGC 4148

seq\_name: gb\_pat:A58686

seq\_documentation\_block:

LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 153)  
Griffiths R. and Tiwari B.

AUTHORS  
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
TITLES  
BIRDS

JOURNAL  
Patent: WO 9639505-A 5 12-DEC-1996;

COMMENT  
ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.

FEATURES  
Location/Qualifiers  
1..153

BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

## alignment\_scores:

Quality: 182.00 Length: 51  
Ratio: 4.439 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 76.471

## alignment\_block:

US-08-973-363-8 x A58686 ..

Align seg 1/1 to: A58686 from: 1 to: 153

1 l1leuProaspProaspPlysPlysProglAlaLysGlnLeuGlnTh 7  
|||||

1 ATTTACCTGATGACCGAGATTAAGAAACACAGCAAGCAAGCTTGACAGAC 50  
8 l1leuProglAlaLysGlnLeuGlnThArgAlaAspTyrLeuIlel 24  
|||||  
51 CAGAAACACAGCAAGCAAGCTTGACAGACCGCGGACATTTACCTCATTA 100  
|||||  
24 l1leuProglAlaLysGlnLeuGlnThArgAlaAspTyrLeuIlel 40  
|||||  
101 AATTACTGAATAAAGACTTGCAAGAAAGAGCAAGCAAGCTTACTGT 150  
|||||  
41 Ala 41  
151 GCA 153

seq\_name: gb\_pat:A58683

seq\_documentation\_block:

LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 153)  
Griffiths R. and Tiwari B.

AUTHORS  
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
TITLES  
BIRDS

JOURNAL  
Patent: WO 9639505-A 2 12-DEC-1996;

COMMENT  
ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.

FEATURES  
Location/Qualifiers  
1..153

BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

## alignment\_scores:

Quality: 172.00 Length: 51  
Ratio: 4.195 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 70.588

## alignment\_block:

US-08-973-363-8 x A58683 ..

Align seg 1/1 to: A58683 from: 1 to: 153

1 l1leuProaspProaspPlysPlysProglAlaLysGlnLeuGlnTh 7  
|||||  
1 ATTTCTCAGATGATCCTGATATAAACACACAAACAAAGCTTACAGAC 50  
|||||  
8 l1leuProglAlaLysGlnLeuGlnThArgAlaAspTyrLeuIlel 24  
|||||  
51 CAAACAAACACAGCAAGCAAGCTTACAGACCGCGGACACTTACTCATTA 100  
|||||  
24 l1leuProglAlaLysGlnLeuGlnThArgAlaAspTyrLeuIlel 40  
|||||  
101 AACTACTGACAGATCTTGCAAGAAAGAGAGGCTCAGACACTTGTGTG 150  
|||||

41 Ala 41  
151 GCG 153

seq\_name: gb\_pr:AF006513

seq\_documentation\_block:

LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997  
DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428





```

17  TATGAAASPTyTleuileuAsnLysAspLeuAlaArgLysG 34
|||||
26882 CCCTGCGACTACCTCATCAATTAAGATAGAGATCTTGCAAAAAAG 26833
34  TAAAGAGT.....CTTCTGTGTCG 26816
|||||
seq_name: gb_pr:AC012624

seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 134365)
DOE Joint Genome Institute.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source 1. 134365
/organism="Homo sapiens"
/adb_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
Quality: 170.00 Length: 41
Ratio: 4.359 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 85.366

alignment_block:
US-08-973-363-8 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365

1 TleuPProAspPProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
118247 ATCTTCATCATGATCCGATTAACCAACCAACAGCTTGACAGAC 118296
17 TATGAAASPTyTleuileuAsnLysAspLeuAlaArgLysG 34
|||||
118297 CCGGCGAGATACCTCATCAATTAAGATAGAGATCTTGCAAAAAAG 118346
34 TAAAGAGT.....CTTCTGTGTCG 118363
118347 AAGCT.....CTTCTGTGTCG 118363
seq_name: gb_htg:AC021449

```

```

seq_documentation_block:
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 143079)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-58M12
JOURNAL Unpublished
2 (bases 1 to 143079)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearrellano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Garayna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczeky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Morrow, J., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
Pierre, N., Pisanl, C., Pollara, Y., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testafaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
JOURNAL Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:7407963.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5154
Center clone name: 58_M12
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134743 bases at least Q40
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 142179; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp

```

```

* 40512 43279: contig of 2768 bp in length
* 43280 43378: gap of 100 bp in length
* 43380 46905: contig of 3528 bp in length
* 46906 47005: gap of 100 bp in length
* 47006 51830: contig of 823 bp in length
* 51831 51926: gap of 100 bp in length
* 51927 62719: contig of 10689 bp in length
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp in length
* 75509 92516: contig of 17008 bp in length
* 92517 92616: gap of 100 bp in length
* 92617 106409: contig of 13793 bp in length
* 106410 106509: gap of 100 bp in length
* 106510 143079: contig of 36570 bp in length.
Location/Qualifiers
1. 143079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC"
misc_feature
1. 38820
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
38921. 40411
/note="assembly-fragment"
misc_feature
40512. 43279
/note="assembly-fragment"
misc_feature
43380. 46905
/note="assembly-fragment"
misc_feature
47006. 51830
/note="assembly-fragment"
misc_feature
51931. 62619
/note="assembly-fragment"
misc_feature
62720. 75408
/note="assembly-fragment"
misc_feature
75509. 92516
/note="assembly-fragment"
misc_feature
92617. 106409
/note="assembly-fragment"
misc_feature
106510. 143079
/note="assembly-fragment"
clone_end:TV
vector_side:right"
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others
ORIGIN
alignment_scores:
Quality: 170.00 Length: 41
Ratio: 4.359 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 85.366
alignment_block:
US-08-973-363-8 x AC021449 ..
Align seg 1/1 to: AC021449 from: 1 to: 143079
1 11leuProaspaspProksplylsProglhAlaLysGlnLeuGlnh 17
|||||
116147 ATTCTCCAGATGATCCGATRAAAAACCAACCAACAGTTGCGAC 116196
|||||
17 rrrgAlasptylleuLleLysleuAnuLysAspLeuAlaTrgLySG 34
|||||
116197 CCGTCACACACCTCATCAATTAATTAGTAGAGCTCTCCAAAAAAG 116246
|||||
34 luAlaGlnAsgLeuAlaGlyAla 41
|||||
116247 AAGCT.....CTTCTGCTGCG 116263
seq_name: gb_htg:AC008531

```

```

seq documentation block:
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.
ACCESSION AC008531
VERSION AC008531.3 GI:12830078
KEYWORDS HTG; HTGS-PHASE2; HTGS-DRAFT.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Substitution
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 14, 2001 this sequence version replaced gi:7528342.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse-field gel estimation
Estimated insert size: 145059; sum-of-contigs estimation
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a working draft sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 56174: contig of 56174 bp in length
* 56175 56274: gap of unknown length
* 56275 100874: contig of 44600 bp in length
* 100875 100974: gap of unknown length
* 100975 113127: contig of 12153 bp in length
* 113128 113228: gap of unknown length
* 113228 118190: contig of 4963 bp in length
* 118191 118290: gap of unknown length
* 118291 118994: contig of 1404 bp in length
* 118995 119794: gap of unknown length
* 119795 123397: contig of 3503 bp in length
* 123398 123498: gap of unknown length
* 123499 145659: contig of 22262 bp in length.
Location/Qualifiers
1. 145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_lib="Caltech human BAC library C"
/clone_lib="CTC-480B11"
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN
alignment_scores:
Quality: 170.00 Length: 41

```

```

Ratio: 4.359                      Gaps: 1
Percent Similarity: 95.122        Percent Identity: 85.366

Alignment block:
US-08-973-363-8 x AC008531    ..

Align seg 1/1 to: AC008531 from: 1 to: 145659

1 TleleuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
ATTCCTCCAGAGCATCCGATAAAAAACCAACAAGCAAAACAGTTCAGAC 46819

17 rrrrglaaspyrleuilelleysteuleuansynsAspLeuAlaArgLysC 34
|||||
CGGTGAGAGCATCCCTCAAAATCTAGTAGAGATCTTGCAAAAAAAG 46869

34 luAlaGlnArgLeuAlaGlyAla 41
|||||
AAAGCT.....CTTCTGGTGGC 46886

seq_name: gb_htg:AC091946

seq_documentation_block:
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 193446)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE
AUTHORS 2 (bases 1 to 193446)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----

Project Information
Center Project Name: 544799
Center clone name: RPCI-11-36012
-----

Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q20
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1358: contig of 1358 bp in length
* 1359 1458: gap of unknown length
* 1459 2667: contig of 1209 bp in length
* 2668 2767: gap of unknown length
* 2768 3823: contig of 1056 bp in length
* 3824 3923: gap of unknown length

```

*	3924	5920:	contig of 1997 bp in length
*	5921	6020:	gap of unknown length
*	6021	7494:	contig of 1474 bp in length
*	7495	7594:	gap of unknown length
*	7595	9096:	contig of 1502 bp in length
*	9097	9196:	gap of unknown length
*	9197	10260:	contig of 1064 bp in length
*	10261	10360:	gap of unknown length
*	10361	12460:	contig of 2100 bp in length
*	12461	12560:	gap of unknown length
*	12561	14611:	contig of 2051 bp in length
*	14612	14711:	gap of unknown length
*	14712	15381:	contig of 1670 bp in length
*	15382	16481:	gap of unknown length
*	16482	17968:	contig of 1487 bp in length
*	17969	18068:	gap of unknown length
*	18069	20434:	contig of 2366 bp in length
*	20435	20534:	gap of unknown length
*	20535	22515:	contig of 2981 bp in length
*	22516	23615:	gap of unknown length
*	23616	27563:	contig of 3448 bp in length
*	27564	30987:	gap of unknown length
*	30988	31087:	contig of 3324 bp in length
*	31088	36065:	contig of unknown length
*	36066	36165:	gap of unknown length
*	36166	40978:	contig of 4813 bp in length
*	40979	41078:	gap of unknown length
*	41079	45663:	contig of 4585 bp in length
*	45664	45763:	gap of unknown length
*	45764	51745:	contig of 5982 bp in length
*	51746	51845:	gap of unknown length
*	51846	57359:	contig of 5514 bp in length
*	57360	57459:	gap of unknown length
*	57460	67881:	contig of 10422 bp in length
*	67882	67981:	gap of unknown length
*	67982	74132:	contig of 6151 bp in length
*	74133	74232:	gap of unknown length
*	74233	79795:	contig of 5563 bp in length
*	79796	79895:	gap of unknown length
*	79896	87511:	contig of 7616 bp in length
*	87512	87611:	gap of unknown length
*	87612	92791:	contig of 5180 bp in length
*	92792	92891:	gap of unknown length
*	92892	102794:	contig of 9903 bp in length
*	102795	102894:	gap of unknown length
*	102895	110866:	contig of 7972 bp in length
*	110867	110966:	gap of unknown length
*	110967	117571:	contig of 6605 bp in length
*	117572	123738:	contig of 6067 bp in length
*	123739	123838:	gap of unknown length
*	123839	130583:	contig of 6745 bp in length
*	130584	130683:	gap of unknown length
*	141545	141544:	contig of 10861 bp in length
*	141645	141644:	gap of unknown length
*	169109:	contig of 27465 bp in length	
*	169110	169209:	gap of unknown length
*	169210	193446:	contig of 24237 bp in length.

FEATURES

source

1..193446

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-360I2"  
/clone\_lib="PC1 human BAC library 11"  
ORIGIN

alignment\_scores:

Quality: 170.00  
Ratio: 4.359

Length: 41  
Gaps: 1

Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:

US-08-973-363-8 x AC091946 ..

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 11leuPProAspAspProAspIysIysProGlnAlaIysGlnLeuGlnth 17  
|||||  
87658 ATTCTCCAGATGATCCGATTAACCAACCAAGCAACAGATTGCGAC 87707

17 rAtgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
87708 CCGTCACACACTACCTCATCAATTACTAGTAGAGATCTTGCAAAAAAAG 87757

34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
87758 AAGCT.....CTTCTGGTGCG 87774

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778

VERSION AC026778.4 GI:14277282

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 195433)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Unpublished

2 (bases 1 to 195433)

DOE Joint Genome Institute.

Direct Submission

Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 195433)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

On Jun 1, 2001 this sequence version replaced gi:13677045.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality >=40 99.9% of Sequence;

Estimated Total Number of Errors is 0.2.

STS Content:

WI-13675 G23101

SHGC-58345 G38487

SHGC-103595 G57841.

FEATURES

source

Location/Qualifiers

1..195433

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

1 11leuPProAspAspProAspIysIysProGlnAlaIysGlnLeuGlnth 17  
|||||  
29732 ATTCTCCAGATGATCCGATTAACCAACCAAGCAACAGATTGCGAC 29683

17 rAtgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
29682 CCGTCACACACTACCTCATCAATTACTAGTAGAGATCTTGCAAAAAAAG 29633

34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
29632 AAGCT.....CTTCTGGTGCG 29616

alignment\_scores:

Quality: 170.00 Length: 41

Ratio: 4.359 Gaps: 1

Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:

US-08-973-363-8 x AC026778/rev ..

---



34 luAlaGlnArgLeuAlaGlyAla 41  
 |||||||||||||||||||||||||||  
 1238 AAGCACAGAGACTGCTGCTGCA 1260

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NM1997.DAT:AAAT42751

seq\_documentation\_block:  
 ID AAAT42751 standard; cDNA: 6608 BP.

AAAT42751;

12-MAR-1997 (first entry)

XX Chicken CHD-1A gene.

DE Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KM CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX Gallus sp.

OS Key Location/Qualifiers  
 FH 228..5390  
 FT CDS /\*tag= a

PN W09639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

PA Griffiths R, Tiwari B;

PI WPI: 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 1; Fig 5; 76pp: English.

XX The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate

CC female development in birds. The sequence of CHD-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AAAT42755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC gene (see also AAAT42756-57). It is located on an autosome or 2

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-rare

CC bird and can be used for sex determ. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores: Quality: 205.00 Length: 41

Ratio: 5.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block: US-08-973-363-8 x AAAT42751 ..

Align seg 1/1 to: AAAT42751 from: 1 to: 6608

1 lIleuPProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
 |||||||||||||||||||||||||||  
 4080 ATTTTACCTGATGATCGACAGAAAGCCAGCAAGAGAGCTACAGAC 4129  
 17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34

|||||||||||||||||||||||||||||  
 4130 CCGTGAGAGACTACCTCATTAATTAAGACTTGCAGAGAAAGG 4179  
 34 luAlaGlnArgLeuAlaGlyAla 41  
 |||||||||||||||||||||||||||  
 4180 AAGCACAAAGCGCTGCTGCTGCA 4202

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NM1997.DAT:AAAT42757

seq\_documentation\_block:  
 ID AAAT42757 standard; DNA: 153 BP.

AAAT42757;

12-MAR-1997 (first entry)

XX Chick CHD-1A gene fragment.

DE Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KM CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

OS Key Location/Qualifiers  
 FH misc\_difference 52..81  
 FT /\*tag= a  
 FT /\*note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"

PN W09639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

PA Griffiths R, Tiwari B;

PI WPI: 1997-043127/04.

XX P-PSDB; AAM08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp: English.

XX Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AAAT42757),

CC chicken CHD-W (W refers to the W chromosome) gene (see also AAAT42758)

CC and the great tit CHD-W gene (see also AAAT42759). Translated amino

CC acid sequences of this region are provided in AAM08146-49. The

CC CHD-1A (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes

CC determine sex in birds and can be used to identify the sex of an

CC embryo, foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores: Quality: 190.00 Length: 51

Ratio: 4.634 Gaps: 1

Percent Similarity: 80.392 Percent Identity: 80.392

alignment\_block: US-08-973-363-8 x AAAT42757 ..

Align seg 1/1 to: AAAT42757 from: 1 to: 153

1 lIleuPProAspAspProAsp..... 7



```

|||||
1 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGACGCTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleI 24
51 CAAGAAACCCAGGCAAGACGCTACAGACCCGCTGACACTTACTCTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AATTACTGATATTAAGACCTTGCAAGAAAGAACACAAAGGCTTGCTGCT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
XX AA142758;
XX
XX 12-MAR-1997 (first entry)
XX
XX Chick CHD-W gene fragment.
XX
XX Bird; sex determination: Chromodomain-helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
XX Gallus sp.
XX
XX Key Location/Qualifiers
XX misc_difference 52..81
XX /*tag= a
XX /note= "bases 52-81 are a repeat of bases 22-51
XX and are ignored in the translated amino
XX acid sequence given in Fig 3"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX PI Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 190.00 Length: 51
Ratio: 4.634 Gaps: 1

```

```

Percent Similarity: 80.392 Percent Identity: 80.392
alignment_block:
us-08-973-363-8 x AA142758 ..
Align seq 1/1 to: AA142758 from: 1 to: 153

1 lleLeuProAspAspProAsp..... 7
|||||
1 ATTTTACCTGATGATCCAGATAAGAAACCCAGGCTAAGACTTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleI 24
51 CAAGAAACCCAGGCTAAGCACTTACAGACCCGCTGACACTTACTCTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AATTACTGATATTAAGACCTTGCAAGAAAGAACAGACAGACTTGCTGCT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759
seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
XX AA142759;
XX
XX 12-MAR-1997 (first entry)
XX
XX Great tit CHD-W gene fragment.
XX
XX Bird; sex determination: chromodomain-helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
XX Parus major.
XX
XX Key Location/Qualifiers
XX misc_difference 52..81
XX /*tag= a
XX /note= "bases 52-81 are a repeat of bases 22-51
XX and are ignored in the translated amino
XX acid sequence given in Fig 3"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX PI Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08149.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,

```

CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;  
 alignment\_scores:  
 Quality: 182.00 Length: 51  
 Ratio: 4.439 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 76.471  
 alignment\_block:  
 US-08-973-363-8 x AAT42759 ..  
 Align seg 1/1 to: AAT42759 from: 1 to: 153  
 1 l1leuProAspProasp..... 7  
 1 ATTTTACCTGATGATCCGATAGAAACACACGCAAGCATTGACAGAC 50  
 8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
 51 CAGAAACACACAGCAAGCAGTTCGACAGCCCGTCGAGATTACTCATTA 100  
 24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
 101 AATTACTGAAATAAGACCTTGCAAGAAAGAGTCAAGACTTACTGT 150  
 41 Ala 41  
 151 GCA 153  
 seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756  
 seq\_documentation\_block:  
 ID AAT42756 standard; DNA; 153 BP.  
 XX AAT42756;  
 AC  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Mouse CHD-1 gene (bases 3855-9777).  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KM CHD-1; CHD-W; W chromosome; ss.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 XX (ISIS-) ISIS INNOVATION LTD.  
 PA  
 XX Griffiths R, Tiwari B;  
 PI  
 XX WPI, 1997-043127/04.  
 DR P-PSDB; AAM08146.  
 XX  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PA Claim 8; Fig 3; 76pp; English.  
 PS

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
 CC and AAT42757). Chicken CHD-W (W refers to the W chromosome) gene  
 CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
 CC Translated amino acid sequences of this region are provided in  
 CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
 CC AAT42754-55) genes determine sex in birds and can be used to identify  
 CC the sex of an embryo, foetus etc. and to manipulate the sex of  
 CC progeny.  
 XX  
 SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
 alignment\_scores:  
 Quality: 172.00 Length: 51  
 Ratio: 4.195 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 70.588  
 alignment\_block:  
 US-08-973-363-8 x AAT42756 ..  
 Align seg 1/1 to: AAT42756 from: 1 to: 153  
 1 l1leuProAspProasp..... 7  
 1 ATTTTACCTGATGATCCGATAGAAACACACAGCAAGCATTGACAGAC 50  
 8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
 51 CAAAGAAACACAGCAAGCAGTTCACAGACCCGTCGAGACTACTCATTA 100  
 24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
 101 AACTACTTAGCAGAGATCTTGCAAGAAAGAGCGTCAAGACTTGTGGT 150  
 41 Ala 41  
 151 GCG 153  
 seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280  
 seq\_documentation\_block:  
 ID AAV59280 standard; cDNA; 1311 BP.  
 XX AAV59280;  
 AC  
 XX  
 DT 14-DEC-1998 (first entry)  
 XX  
 DE Altered telomere repeat binding factor 1 gene.  
 XX  
 KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 KM telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
 XX  
 OS Homo sapiens.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1311  
 FT /tag= a  
 FT /product= "A-TRF"  
 XX  
 PN WO9836066-A1.  
 XX  
 PD 20-AUG-1998.  
 XX  
 PF 13-FEB-1998; 98WO-US02765.  
 XX  
 PR 04-FEB-1998; 98US-0018628.  
 PR 13-FEB-1997; 97US-0800264.  
 XX  
 XX (U9KQ ) UNIV ROCKEFELLER.  
 XX

PI Bianchi A, De Lange T, Van Steensel B;  
XX WPI: 1998-480769/41.  
DR P-PSDB; AAW59280.  
XX  
PT Nucleic acid encoding altered telomere repeat binding protein and  
PT related vectors - transformants, hetero-dimers and antibodies, used  
PT to inhibit shortening of telomerases caused by ageing or disease.  
XX also used to extend life of cells in culture  
XX  
PS Claim 14, Page 110-111, 163pp; English.  
XX  
CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
CC telomere repeat binding factor (TRF) dimerization domain, and forms a  
CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
CC inhibit shortening of telomeres associated with ageing (for cosmetic  
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
CC tumours and viral (including human immune deficiency virus) infection.  
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
CC for expression of recombinant proteins or where intended for subsequent  
CC transplant or for testing, eliminating the need for transformation.  
XX  
SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;  
  
alignment\_scores:  
Quality: 129.00 Length: 25  
Ratio: 5.160 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-8 x AAW59280 ..  
  
Align seg 1/1 to: AAW59280 from: 1 to: 1311  
  
11leleuproaspasproasplyslysglysproglnalalysglnleuglnrh 17  
1237 ATTCTCCAGATGATCGTATATAAAACACACACAAACAGTACAGAC 1286  
17 rarglaasptyrleuilelyslsu 25  
1287 CCGTCAGACTACCTCATCAAACTA 1311  
  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA2001A.DAT.AAK88882  
seq\_documentation\_block:  
ID AAK88882 standard; cDNA; 421 BP.  
XX  
AC AAK88882;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
XX  
XX Human, digestive system antigen; gene therapy; cancer; appendicitis;  
KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KM digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229345.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.

CC	XX	PA	13-OCT-2000	2000US-0239937	
CC	XX	PA	20-OCT-2000	2000US-0249960	
CC	XX	PA	20-OCT-2000	2000US-0241221	
CC	XX	PA	20-OCT-2000	2000US-0241785	
CC	XX	PA	20-OCT-2000	2000US-0241786	
CC	XX	PA	20-OCT-2000	2000US-0241787	
CC	XX	PA	20-OCT-2000	2000US-0241808	
CC	XX	PA	20-OCT-2000	2000US-0241809	
CC	XX	PA	20-OCT-2000	2000US-0241826	
CC	XX	PA	01-NOV-2000	2000US-0244617	
CC	XX	PA	08-NOV-2000	2000US-0246474	
CC	XX	PA	08-NOV-2000	2000US-0246475	
CC	XX	PA	08-NOV-2000	2000US-0246476	
CC	XX	PA	08-NOV-2000	2000US-0246477	
CC	XX	PA	08-NOV-2000	2000US-0246478	
CC	XX	PA	08-NOV-2000	2000US-0246523	
CC	XX	PA	08-NOV-2000	2000US-0246524	
CC	XX	PA	08-NOV-2000	2000US-0246525	
CC	XX	PA	08-NOV-2000	2000US-0246526	
CC	XX	PA	08-NOV-2000	2000US-0246527	
CC	XX	PA	08-NOV-2000	2000US-0246528	
CC	XX	PA	08-NOV-2000	2000US-0246532	
CC	XX	PA	08-NOV-2000	2000US-0246609	
CC	XX	PA	08-NOV-2000	2000US-0246610	
CC	XX	PA	08-NOV-2000	2000US-0246611	
CC	XX	PA	08-NOV-2000	2000US-0246613	
CC	XX	PA	17-NOV-2000	2000US-0249207	
CC	XX	PA	17-NOV-2000	2000US-0249208	
CC	XX	PA	17-NOV-2000	2000US-0249209	
CC	XX	PA	17-NOV-2000	2000US-0249210	
CC	XX	PA	17-NOV-2000	2000US-0249211	
CC	XX	PA	17-NOV-2000	2000US-0249212	
CC	XX	PA	17-NOV-2000	2000US-0249213	
CC	XX	PA	17-NOV-2000	2000US-0249214	
CC	XX	PA	17-NOV-2000	2000US-0249215	
CC	XX	PA	17-NOV-2000	2000US-0249216	
CC	XX	PA	17-NOV-2000	2000US-0249217	
CC	XX	PA	17-NOV-2000	2000US-0249218	
CC	XX	PA	17-NOV-2000	2000US-0249245	
CC	XX	PA	17-NOV-2000	2000US-0249245	
CC	XX	PA	17-NOV-2000	2000US-0249264	
CC	XX	PA	17-NOV-2000	2000US-0249265	
CC	XX	PA	17-NOV-2000	2000US-0249297	
CC	XX	PA	17-NOV-2000	2000US-0249299	
CC	XX	PA	17-NOV-2000	2000US-0249300	
CC	XX	PA	01-DEC-2000	2000US-0250160	
CC	XX	PA	01-DEC-2000	2000US-0250391	
CC	XX	PA	05-DEC-2000	2000US-0251030	
CC	XX	PA	05-DEC-2000	2000US-0251038	
CC	XX	PA	05-DEC-2000	2000US-0256719	
CC	XX	PA	06-DEC-2000	2000US-0251479	
CC	XX	PA	08-DEC-2000	2000US-0251856	
CC	XX	PA	08-DEC-2000	2000US-0251868	
CC	XX	PA	08-DEC-2000	2000US-0251869	
CC	XX	PA	08-DEC-2000	2000US-0251989	
CC	XX	PA	08-DEC-2000	2000US-0251990	
CC	XX	PA	11-DEC-2000	2000US-0254097	
CC	XX	PA	03-JAN-2001	2001US-0259678	
CC	XX	PA	(HUMA-) HUMAN GENOME SCI INC.		
CC	XX	PI	Rosen CA, Barash SC, Ruben SM;		
CC	XX	DR	WPI: 2001-502630/55.		
CC	XX	DR	P-PSDB: AAM93109.		
CC	XX	PT	Polynucleotides encoding digestive antigens, useful for		
CC	XX	PT	diagnosing, treating, preventing and/or prognosing disorders of the		
CC	XX	PT	digestive system, particularly cancer and cancer metastases -		
CC	XX	PS	Claim 1; SEQ ID NO 1198; 986bp; English.		
CC	XX	CC	The present invention provides the protein and coding sequences of a		

```

CC number if human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC system antigen of the invention.
CC
SO Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other:

alignment_scores:
    Quality: 122.00      Length: 35
    Ratio: 4.067         Gaps: 0
    Percent Similarity: 85.714      Percent Identity: 71.429

alignment_block:
US-08-973-363-8 x AAK88882 ..

Align seg 1/1 to: AAK88882 from: 1 to: 421

1 HleleuProbsapProbsapLysProGlnAlaLysGlnLeuGlnH 17
|||||.....:|||||.....:|||||.....:|||||
110 ATTCTGCGGTGGAGACGATTAAGCCCTCAGGGGAGAGCCTACGAC 159
17 rAAGAAsPtyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34
|||||.....:|||||.....:|||||.....:|||||
160 CCGAGCGGAGTACTTGTGGAAGCTGCTCAGAAAGGGTCTGAGAGAGAG 209
34 LuAla 35
|||
210 GGGCT 214

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA157603
seq_documentation_block:
ID AA157603 standard; cDNA; 421 BP.
XX
AC AA157603;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.

```



Percent Similarity: 85.714 Percent Identity: 71.429

alignment\_block:  
US-08-973-363-8 x AAI57603 ..

Align seg 1/1 to: AAI57603 from: 1 to: 421

```
1 11leuProAspAspProAspLysProGlnAlaLysGlnLeuGlnrh 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
110 ATTCTGCGGTGGAGACAGATATAAAAGCCTCAGGGGAAACACTCTACAGC 159
17 rargAlaAspTyrLeuLeuLeuLysLeuLysAspLeuAlaArgL 34
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
160 CCGACCGGATCTTCTGTGACAGCTCTCAGAAAGGCTCTGAGAGAAAGG 209
34 LuAla 35
|||
210 GGGCT 214
```

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:

ID ABL06443 standard; cDNA; 6240 BP.

```
XX AC ABL06443:
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB62340.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;
```

alignment\_scores:

Quality: 88.00 Length: 38  
Ratio: 2.839 Gaps: 2  
Percent Similarity: 81.579 Percent Identity: 55.263

alignment\_block:  
US-08-973-363-8 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```
1 11leuProAspAspProAspLysProGlnAlaLysGlnLeuGlnrh 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4184 ATTTCCTTAACGAT...ACGGCAAGCCCCAGCCAAACACTCTCAGAC 4230
17 rargAlaAspTyrLeuLeuLeuLysLeuLysAspLeu...AlaArgL 33
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4231 GCGTCCGAGTACTCTCTCAGATCATCAAGAAAGACGTGAGCTGACCA 4280
33 ysgLuAlaGlnArg 37
|| :|||
4281 ACGGAGCACACGCG 4294
```

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:

ID ABL06442 standard; cDNA; 9933 BP.

```
XX AC ABL06442:
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB62339.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;
```

alignment\_scores:                   Quality: 88.00                   Length: 38  
                                   Ratio: 2.839                   Gaps: 2  
                                   Percent Similarity: 81.579           Percent Identity: 55.263

alignment\_block:  
 US-08-973-363-8 x AB106442 ..

Align seg 1/1 to: AB106442 from: 1 to: 9933

1   1leuPProAspAProAspLysLysProGlnAlaLysGlnLeuGlnth 17  
       |||||                   |||||                   |||||  
       6724 ATTTGGCTTAACGAT...ACGCCAACGCCCCAGCCAGCAGCTGCACGAC 6770

17   rArGAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeu...AlaArgL 33  
       |||||                   |||||                   |||||  
       6771 GCGTGCAGCTACCTGCTCAGATCATCATGAGAAAGACGTGAGCTGACCA 6820

33   ysGluAlaGlnArg 37  
       ||                   |||||  
       6821 AGGAGAGACAAACGC 6834

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AA559590

seq\_documentation\_block:  
 ID   AA559590 standard: DNA: 12928 BP.  
 AC   AA559590;  
 XX     
 DT   13-FEB-2002 (first entry)  
 XX     
 DE   Propionibacterium acnes immunogenic protein encoding DNA #85.  
 XX     
 KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW   dermatological; osteopathic; neuroprotectant; ds.  
 OS   Propionibacterium acnes.  
 XX     
 PN   W0200181581-A2.  
 XX     
 PD   01-NOV-2001.  
 XX     
 PE   20-APR-2001; 2001WO-US12865.  
 XX     
 PR   21-APR-2000; 2000US-199047P.  
 PR   02-JUN-2000; 2000US-208841P.  
 PR   07-JUL-2000; 2000US-216747P.  
 XX     
 PA   (CORI-) CORIXA CORP.  
 XX     
 PI   Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI   L'oiseleur J, Zhang Y, Jen S, Carter D;  
 DR   WPI: 2001-616774/71.  
 XX     
 PT   Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT   vaccinating against and diagnosing infections, especially useful for  
 PT   treating acne vulgaris -  
 XX     
 PS   Claim 1: SEQ ID No 85; 1069PP; English.  
 CC   Sequences AA559506-AA559804 represent DNA molecules encoding  
 CC   Propionibacterium acnes immunogenic polypeptides. The proteins and their  
 CC   associated DNA sequences are used in the treatment, prevention and  
 CC   diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC   SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
 CC   osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC   in infections of bone, joints and the central nervous system, however it  
 CC   is particularly involved in the inflammatory lesions associated with acne  
 CC   vulgaris. A method for detecting the presence or absence of P. acnes in a

CC   patient comprises contacting a sample with a binding agent that binds to  
 CC   the proteins of the invention and determining the amount of bound protein  
 CC   in the sample. The polypeptides may be used as antigens in the production  
 CC   of antibodies specific for P. acnes proteins. These antibodies can be  
 CC   used to downregulate expression and activity of P. acnes polypeptides and  
 CC   therefore treat P. acnes infections. The antibodies may also be used as  
 CC   diagnostic agents for determining P. acnes presence, for example, by  
 CC   enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC   polypeptides shown in AAU58255-AAU58404 and AAU67601-AAU67602.  
 CC   Note: The sequence data for this patent did not form part of the printed  
 CC   specification, but was obtained in electronic format directly from WIPO  
 CC   at ftp.wipo.int/pub/published\_pot\_sequences.  
 XX     
 SQ   Sequence 12928 BP: 2564 A; 4099 C; 4037 G; 2225 T; 3 other;

alignment\_scores:                   Quality: 61.00                   Length: 40  
                                   Ratio: 2.103                   Gaps: 1  
                                   Percent Similarity: 72.500           Percent Identity: 32.500

alignment\_block:  
 US-08-973-363-8 x AA559590 ..

Align seg 1/1 to: AA559590 from: 1 to: 12928

3   ProAspAProAspLysLysProGlnAlaLysGln.....LeuGlnth 17  
       |||||                   |||||                   |||||  
       7295 CCGGACGCTCATCCGCCACCCCGCAAGCCGCAAGACGCTACTTCCAC 7344

17   rArGAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
       ||                   |||||                   |||||  
       7345 GTCATCGACGCGTACTCGACGATCTCGAAGAAAGATGTTCCCTGACCA 7394

34   luAlaGlnArgLeuAlaGly 40  
       |||||                   |||||  
       7395 CTGGAGACGCTGTAGCGCGGA 7414

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AA596071

seq\_documentation\_block:  
 ID   AA596071 standard: cDNA: 1908 BP.  
 AC   AA596071;  
 XX     
 DT   26-FEB-2002 (first entry)  
 XX     
 DE   Leishmania antigen 4G2-83 extended DNA.  
 XX     
 KW   Leishmaniasis; Leishmania antigen; immunostimulant; protozoacide; Ldp23;  
 KW   interleukin-15; PCR primer; Lbhp83; M15; Lt-1; LbeIF4A; LmgSP9a;  
 KW   MAPS-1A; LmgSP1; LmgSP3; LmgSP5; LmgSP8; LmgSP9; LmgSP13; LmgSP19; ss;  
 KW   LcgSP1; LcgSP3; LcgSP4; LcgSP8; LcgSP10; 1G6-34; 1B6-44; 4A5-63; 1B11-39;  
 KW   2A10-37; 4G2-83; 4H6-41; 8G3-100.  
 XX     
 OS   Leishmania major.  
 XX     
 PN   W0200179276-A2.  
 XX     
 PD   25-OCT-2001.  
 XX     
 PE   05-APR-2001; 2001WO-US11254.  
 XX     
 PR   14-APR-2000; 2000US-0551974.  
 PR   05-MAY-2000; 2000US-0565501.  
 PR   14-AUG-2000; 2000US-0639206.  
 XX     
 PA   (CORI-) CORIXA CORP.  
 XX     
 PI   Reed SG, Campos-Neto A, Webb JR, Dillon DC, Skelky YAM, Bhatia A;  
 PI   Coker RM, Probst P;  
 DR   WPI: 2002-061971/08.







**THIS PAGE BLANK (USPTO)**



```

seq.documentation_block:
LOCUS          BBI155356          619 bp      mRNA      linear      EST 18-OCT-2001
DEFINITION     BBI155356 RIKEN full-length enriched; 16 days neonate thymus Mus
                musculus cDNA clone A130024L16.3 similar to L10410 Mouse
                DNA-binding protein (CHD-1) mRNA, mRNA sequence.
ACCESSION      BBI155356
VERSION        BBI155356.2 GI:16268254
KEYWORDS       EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 619)
AUTHORS        Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
                Himoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda
                M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
                Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
                D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
                Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Toyota,T.,
                Muramatsu,M. and Hayashizaki,Y.
                RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
TITLE          On Jun 29, 2000 this sequence version replaced gi:1811286.
JOURNAL        Unpublished (2001)
COMMENT        Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center (GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,
                URL: http://genome.gsc.riken.go.jp/
                Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
                M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new
                genes. Genome Res. 10 (10) 1617-1630 (2000)
                wagui,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
                Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
                S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
                Hayashizaki,Y.
                RIKEN integrated sequence analysis (RISA) system-384-format
                sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                10 (11), 1757-1771 (2000)
                Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
                Y. and Hayashizaki,Y.
                Computer-based methods for the mouse full-length cDNA
                encyclopedia: real-time sequence clustering for construction of a
                nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Alizawa
                K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
                Hayashizaki,Y.
                Computational Analysis of Full-length Mouse cDNAs Compared with
                Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
                Please visit our web site (http://genome.gsc.riken.go.jp/) for
                further details.
                cDNA library was prepared and sequenced in Mouse Genome
                Encyclopedia Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in RIKEN.
                Division of Experimental Animal Research in Riken contributed to
                prepare mouse tissues.
FEATURES
             source
             1..619
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /clone="A130024L16"
             /clone_id="RIKEN full-length enriched, 16 days neonate
             thymus"
             /tissue_type="thymus"
             /dev_stage="16 days neonate"
             /lab_host="DH10B"
             /note="Site_1: Sali, Site_2: BamHI; cDNA library was
             prepared and sequenced in Mouse Genome Encyclopedia

```

```

Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAAGATCCACAGACTCTTTTGTTCCTTTTTTNNV 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTTCAGATTAATAAATTAATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pbluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT      194 a       118 c       161 g       146 t
ORIGIN

alignment_scores:
    Quality:   187.00          Length:     41
    Ratio:     4.561           Gaps:        0
Percent Similarity: 100.000    Percent Identity: 87.805

alignment_block:
US-08-973-363-8 x BB155356 ..

Align seg 1/1 to: BB155356 from: 1 to: 619

1 LIELEUPRICASPAPSPROASPLYSLSYSGPROGLIALILYSGLINLEUGLNTH 17
|||||.....
489 ATTCTTCGCAGATGATCTCATATANAACCACAAGCAAAACAGTTACAGAC 538

17 TARGALAAAPTYYLLEULIELLYSLDEUNASNYLASPLEUALAARYLYSG 34
|||||.....
539 CGGTGCAGACTACCTCATCACTACTTTCAGAGAGATCTTGCAAAAAGAG 588

34 LUUAGLUARGLUUAIGLYALA 41
|||||.....
589 AGGCTCGAGAGACTTGTGGTGGC 611

seq_name= gb_estl:AL644594

seq_documentation_block:
LOCUS      AL644594      645 bp      mRNA      linear      EST 07-NOV-2001
DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone L1EId12 5', mRNA
sequence.
ACCESSION  AF644594
VERSION    AF644594.1 GI:16796719
KEYWORDS   EST.
SOURCE     western clawed frog.
ORGANISM   Silurana tropicalis
            Euryptora; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
            Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 645)
AUTHORS   Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE      Sanger xenopus tropicalis EST project 2001 (10_2001)
JOURNAL   Unpublished (2001)
COMMENT    Contact: Huckle E
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE.ID: L1EId12.plc
            Sequencing primer: PLC
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
            Location/Qualifiers
                1..645
                   /organism="Silurana tropicalis"
                   /db_xref="taxon:8364"
                   /clone="L1EId12"
FEATURES
source

```

```

/clove, lib="XCC-seg"
/dev_stage="egg"
//host="Escherichia coli XL1-blue"
//note="vector: pCS107; site.1: EcoRI; site.2: NotI; cDNA
was oligo dT primed from 5' end of poly A+ RNA from egg
embryo-notI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

BASE COUNT      222 a      125 c      156 g      141 t      1 others

ORIGIN

Alignment Scores:
      Quality: 187.00      Length: 39
      Ratio: 4.921      Gaps: 0
Percent Similarity: 97.436      Percent Identity: 94.872

Alignment block:
US-08-973-363-8 x AL644594

Align seg 1/1 to: AL644594 from: 1 to: 645

1 11elePuoAspAspProAspLyLySProGlnAlaLySgInleuGlnth 17
|||||
456 ATTATACCGATGATGCAGTAAGAACCCCAAGCAAGAGCTTAAGAC 505
17 TATGATaAspTyrLeu11eLySLeuAsnLySAspLeuAlaLySg 34
|||||
506 CAGAGTCGACTACCTCATTTAACTTCTCAATTAAGATCTGTTAGGAAG 555

TITLE
34 lvalaglnarlguala 39
|||||
556 AAGCACAAAGACTTCT 572

seq_name: gb_est1:BB461065

seq documentation block:
LOCUS      BB461065              660 bp      mRNA      linear      EST 25-OCT-2001
DEFINITION      BB461065 RIKEN full-length enriched, 12 days embryonic spinal ganglion
Mus musculus cDNA clone D130070b13.3 similar to L10410 Mouse
DNA-binding protein (Cbp-1) mRNA, mRNA sequence.
ACCESSION      BB461065
VERSION        BB461065.2      GI:16426612
KEYWORDS
SOURCE        house mouse.
ORGANISM      Mus musculus
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 660)
AATkAACTC.T, Carinini.P., Fukuda.S., Furuno.M., Hanagaki.T., Hara.A,
Aikawa.T., Kori.F., Ishii.Y., Ito.M., Kawai.J., Konno.H., Kouda.A.,
M., Koya.S., Matsuyama.T., Miyazaki.A., Nomura.K., Ohno.M.,
Okazaki.Y., Okito.T., Saito.R., Sakai.C., Sakai.K., Sano.H., Sasaaki.H.,
D., Shibata.K., Shinagawa.A., Shitaki.T., Sogabe.Y., Suzuki.H.,
Tegami.M., Tagawa.A., Takahashi.T., Takeda.Y., Tanaka.T., Toya.T.,
Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Arikawa.T., et al. 2001)
Unpublished (2001)
On Jul 21, 2000 this sequence version replaced gi:9356558.
Contact: Yoshinobu Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suifu-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carinini.P., Shibata.Y., Hayashi.N., Sugahara.Y., Shibata.K., Itoh
M., Kono.H., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
vgli.K., Fujiwake.S., Inoue.K., Togawa.Y., Izawa.M., Ohara.E.,

```

```

MATCHIKI,M., YONEDA,Y., ISHIKAWA,T., OZAWA,K., TANAKA,T., MATSUURU
,S., KAWAI,Y., OKAZAKI,Y., MURAMATSU,M., INOUE,Y., KITAH,A. and
Hayashizaki,Y.
Riken Integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1657-1671 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Garnincel,P., Sugahara
,Y. and Hayashizaki,Y.
Computer based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kitayama,H., Yamamaka,I.,
,K., Fukuda,S., Hata,A.A., Itoh,M., Kawai,Y., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1..660
/oranism="Mus musculus"
/db_xref="taxon:10090"
/cclone."D13007B13"
/clone.lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH108"
/note="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAAGACGGCCCGCACTGCAGCTTTTTCCTTTTNN 3'], cDNA was
prepared by using triethanolamine thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAAGACTCCTGCAGCTTAATTAATAATTAATCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
phluescript KS(+) after bulk excision from lambda FLIC I."
BASE COUNT 211 a 130 c 168 g 151 t
ORIGIN
alignment_scores:
Quality: 187.00 Length: 41
Ratio: 4.561 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.805
alignment_block:
US-08-973-363-8 x BB461065 ..
Align seg 1/1 to: BB461065 from: 1 to: 660
1 IIleleProaspaspproaspilyslyssproginalytsgintlenclnht 17
|||||.....
481 ATTCTTCCCATGTATCTCTGATATAAAAAACCAACGAACAAGCTTAAAGC 530
rAtgAlaAspyrTleutlleTlysleuAsnTylAspsleuAlaArgLsg 34
|||||.....
531 CGGTGGAGACACTCTCATCAACACTCTATTACAGACAGATCTTGCAAAGAAG 580
34 luAaGlnArGLeuAlaGyaLa 41
|||||.....
581 AGGCTTAGAGACTTGGTGGTGG 603

```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:  
LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930033021 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922.1 GI:17013165  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 446)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
Y., Ito,M., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse full-length cDNAs (Akimura,T., et al.  
2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp.  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

FEATURES  
source  
1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033021"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores: Length: 41  
Quality: 182.00  
Ratio: 4.439 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 82.927

alignment\_block:  
US-08-973-363-8 x BB834922 ..

Align seg 1/1 to: BB834922 from: 1 to: 446

1 l l l e u d p r o a s p p r o a s p l y s p r o g l n a l a y s c l n l e u g l n t h 17  
|||||  
309 A T T C T T C C A G A T G A T C C T G A T A A A A A C C A A G C C A A G C T T A C A G 358  
17 r a r g a l a a s p t y r l e u l l e y s t e u l e u a s n l y a s p l e u a l a r g y s g 34  
|||||  
359 C C G T C A G A C T A C C T C C T C A A A C T A C G A G A G A C T T G C C A A A G A G 408  
34 l u a l a g l n a r g l e u a l a g l a l y a l a 41  
|||||  
409 A G G C T C A G A G A C T T G T G G T T C G 431

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:  
LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730.1 GI:17008973  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 438)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
Y., Ito,M., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse full-length cDNAs (Akimura,T., et al.  
2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp.  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.



Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012" 114 a 118 c 85 g 230 t

BASE COUNT  
ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-8 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

111eleupProaspPpRoaspPlylsySProglInAlalysglInuenglnth 17  
|||||  
378 ATTCTTCAGATGATCCGATTAATAAACCAACAAACAGTTGCAGAC 329  
17 rArgAlaAspTYrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
328 CGGTGCAGACTACCTCATCAATTACTTAGTAGAGATCTTGCAAAAAAG 279  
34 luAglInArGlEuAlaGlYAla 41  
|||||  
278 AAGCT.....CTTCTGTGTCG 262

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isogai,T.)  
Unpublished (2000)  
JOURNAL  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
Location/Qualifiers  
1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-8 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

111eleupProaspPpRoaspPlylsySProglInAlalysglInuenglnth 17  
|||||  
450 ATTCTTCAGATGATCCGATTAATAAACCAACAAACAGTTGCAGAC 499  
17 rArgAlaAspTYrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
500 CGGTGCAGACTACCTCATCAATTACTTAGTAGAGATCTTGCAAAAAAG 549  
34 luAglInArGlEuAlaGlYAla 41  
|||||  
550 AAGCT.....CTTCTGTGTCG 566

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:  
LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060P1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1028)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DFP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>  
Plate: LLM9753 row: h column: 16  
High quality sequence stop: 488.  
FEATURES  
Location/Qualifiers  
1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH-MGC\_72"  
/clone\_type="melanotic melanoma"  
/tissue="organ: skin; Vector: pcMV-SPORE6; Site:1; NotI;  
Site:2; SalI; Cloned unidirectionally. Primer: oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-8 x AU125712 ..



US-08-973-363-8 x BE895133

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 lleauproaspaprosaplyslsproglnalalysglnleuglnphr 17  
 |||||  
 131 ATTCTTCACATGATCCGATTAATAAAACCAACCAACAGCTTGACAGC 180  
 17 rarglaasptyleuilelyslleuenaenylsaspleuialargylg 34  
 |||||  
 181 CCGTCACTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 230  
 34 luialaglnargleuialaglyala 41  
 |||||  
 231 AAGCT.....CTTCTGCTGCG 247

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 leauproaspaprosaplyslsproglnalalysglnleuglnphr 18  
 |||||  
 674 CCGCCGATATCTCTATTAAGACCTCAGGCCAAGCAAGCTTACAGACG 625  
 18 glnlaasptyleuilelyslleuenaenylsaspleuialargylg 35  
 |||||  
 624 AGCCACTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 575  
 35 laglnargleuialagly 40  
 |||||  
 574 CCCACACACAGGSGGCG 558

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

Page 7

Align seg 1/1 to: BF239967 from: 1 to: 821

6 ProAspLysVSProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22  
 |||||  
 3 CCGCATAAAAAACCAAGCAAAACAGTGTGCAGACCCGTCGACACTACT 52  
 22 uUleLysLeuLeuAsnLysAspLeuAlaArgTysGlnAlaGlnArgLeu 39  
 |||||  
 53 CATCAATTACTTACTAGTACAGATCTTGCAAAAAGAGAGCT.....CTTT 96  
 39 IaGlyAla 41  
 :|||  
 97 CTGTGTGG 104

```
seq_name: gb_est1:AW996787
seq_documentation_block:
```

BASE COUNT	83 a	59 c	72 g	123 t
ORIGIN				

alignment_scores:		
Quality:	133.50	Length: 36
Ratio:	3.926	Gaps: 2
Percent Similarity:	94.444	Percent Identity: 83.333

```
alignment_block;
```

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

1 lleuaprkaspasprrpaaaplyslvysppogialalalygslmeuailrh 17  
105 attcttccatcatgattcccatataaaaacccacacaaacaaacagttgcagac 56  
17 ratgalaasptryleuailelys.leuileuansnlysaspleualatglys 33  
55 ccgagacagactatctcatcatcaatcctt...agtagaactcttgcaaaaaa 9

seq_name:	gb_estl:AM997058					
seq_documentation_block:						
LOCUS	AM997058	686	bp	mRNA	linear	EST 05-JUN-2000
DEFINITION	QW7-BND0047-150400-152-c03	BND047	Homo sapiens	CDNA,	mRNA	sequence.
ACCESSION	AM997058					
VERSION	AM997058.1	GI:8257292				
KEYWORDS	EST.					
SOURCE	human,					

BASE COUNT	165 a	154 c	126 g	241 t
ORIGIN				

## alignment\_scores:

Quality: 127.00 Length: 35  
Ratio: 3.629 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 82.857

## alignment\_block:

US-08-973-363-8 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

1 lileuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
127 ATTCTTCAGATGATCCGATTA AAAACCAAGCAAAACGTTCCAGAC 79  
|||||  
17 rArgAlaSPtyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
78 CCGTCAGACTACCTCATCAATTACTTAGAGAGATCTCCAAAAAGAA 29  
34 luAla 35  
|||||  
28 GCTCT 24

seq\_name: gb\_gss:CNS05T9J

seq\_documentation\_block:

LOCUS CNS05T9J 1122 bp DNA linear GSS 26-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
042M09 of library C from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL352864.1 GI:8246657  
VERSION AL352864  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis.

REFERENCE 1 (bases 1 to 1122)  
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished

JOURNAL 2 (bases 1 to 1122)  
REFERENCE Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,  
AUTHORS Bernot,A., Fizames,C., Mincker,P., Brotlier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished

JOURNAL 3 (bases 1 to 1122)  
REFERENCE Genoscope  
AUTHORS Direct Submission  
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

FEATURES  
Source  
1..1122  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="042M09"  
/clone\_lib="C"  
/note="Genoscope sequence ID : C0BC042AG05C1-end : T7"

BASE COUNT 303 a 260 c 301 g 246 t 12 others  
ORIGIN

## alignment\_scores:

Quality: 124.00

Length: 29

Ratio: 4.276 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 79.310

## alignment\_block:

US-08-973-363-8 x CNS05T9J ..

Align seg 1/1 to: CNS05T9J from: 1 to: 1122

1 lileuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
835 ATTCTACCAAGATGATCCAGCAAGAGCCCGAGTCCAGCAGCTACAGGC 884  
|||||  
17 rArgAlaSPtyrLeuIleLysLeuAsnLysAsp 29  
|||||  
885 CAGAGCTGAGTATCTCTCCTCAAGCTGCTGAAGAAAGGAC 921





```

source
1. .6608
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2289 a 1207 c 1459 g 1653 t
ORIGIN

alignment_scores:
Quality: 197.00 Length: 41
Ratio: 4.805 Gaps: 0
Percent similarity: 100.000 Percent identity: 95.122

alignment_block:
US-08-973-363-9 x A58691 ..

Align seg 1/1 to: A58691 from: 1 to: 6608

1 IIELEUPROASPASPAPROASPLYSIVSProGlnAlaIysGlnLeuGlnth 17
|||||
4080 ATTATTACCTGATGATCCACAGACAAACCCCGAGCAACAGCTTACAGAC 4129
|||||
17 rARGAlaAsPtyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34
|||||
4130 CCGTGAGAGACTACCTCATTAATTTACTGATATGAATAAGACCTTGCAAGAAG 4179
|||||
34 luvAlGlnArgLeuthrGlyAla 41
|||||
4180 AAGCACAAAGGCTTGCTGCTGCA 4202

seq_name: gb_ov:AF004397

seq_documentation_block:
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
cds.
ACCESSION AF004397
VERSION AF004397.1 GI:2501845
SOURCE
ORGANISM
.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 6872)
Griffiths, R. and Korn, R.M.
A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus
1974/73516
2 (bases 1 to 6872)
Griffiths, R. and Korn, R.M.
Direct Submission
Submitted (16-MAY-1997) ZOOLOGY, Molecular Lab, Glasgow University,
Glasgow G12 8QQ, UK

FEATURES
source
Location/Qualifiers
1..6872
/organism="Gallus gallus"
/db_xref="taxon:9031"
1..6872
/feature="CHD-Z"
228..5654
/feature="CHD-Z"
/function="role in chromatin architecture"
/note="CHD protein with hydrophilic domain"
/codon_start=1
/product="chromo-helicase-DNA-binding on the Z chromosome
protein"
/protein_id="AAC60282.1"
/db_xref="GI:2501846"
/translation="MNGHSDSEYRNSSGSRSDSDSGASGSGSSSGSSSDGSS
SQSSSSSESGESGSGSESSESDTSRKKQYQAPRADSGEPFKSSPSITLAVRSAY
LKQKQDQKAASDSGSEEDSSSDSDADSSSETKRKHKHEDWQSGSVSGTSS
DSEAEQGDGKSSCESESDYEPKKKVSRRKPPSRITPKSGKKSTGQKRLDSEEEBEE

```

DDDDVYKRGSRRAATVNVSAKAEELTKTOSDLLEVCQEDVDPVTEDEEPTIEKPFK  
SRICRKRATGASTITIVAEVADGDPNNAFEKSEKLEGEIYOLCITKMGNSHINHTWTEET  
LKQNVAGNMKLDWTKKIDDETDGRKMLNASPEDEYIYNQDELTDLHKOIÖIVERITII  
AHSNQSAAGAPDYCCMOGLPYSECSMEGALIAKKEFOAIDEFYRNOSKTPFDK  
CXYLQKPREVALKQPSYIGHESLELRQYLQGLNMWLAHSMCGKNSCIADMGGLG  
KTIOITSEFLNLFIEHOLYGEFLLRVPLSTWOREIGTMAPQMAVYVYIGDITSRN  
MIRHEMMHDPOTKRILKNILITTYEILLKKSPLGJNMAFTGYDEAHLKNDSDILXX  
RTLIDFNSNRBLITTPPLONSLEKELMSLHFIIMPEKFSWDEEHEBKGREGYKAS  
LHELEFPFLLRVKKDVEKSLPAKVEOILREMSALOKÖYKWLITRNKALSGSKG  
STSEFLIMMELKKCNHCYLIKPPDNEFENKOEALÖHLIRSSGLILDLKLIRL  
ERGSRVYLIFSQMRMLDILAEYLKRRPFQRLDGSYIKGELRKQALDPFNAEGSEDFC  
FLSTRSGGGINLASADPVYIPDSNPNPODLQOARHIGQKOVNIYRLVTKGS  
VEEDILERRAKKMWLDHLYIORMOTTKTYLHGSPSSPPPEKKEPLSATIKRGAE  
LFPEBEEDEBPQMDIDELIKRAETENEBGLTYVDELLSQRKANFSMDDDIIE  
LEPERNSRNMEEIIPESQRRRIEEERÖKLELEETVLPMPMNCAKÖISFNGSEGRSR  
SRRSGSDSISITERKPKRKGRPERTPREMIKGFSAELIRFITSYKKFGEPRLD  
AVADALVJKSESTDRLRLELGVNCGIKALKDSSGÖERAGLGYKGFPTISGRI  
ÖVANKATISHHEELAPLHKISIPDSPEERKRVYICHHKRAHFIDIDMCKEDDSNLVG  
YEGYGSWEMIKMPPDLSIÖKIPDPDPKPKÖKÖIÖATRDYILKINDLAKKEKÖ  
RLGAGNSKRTKTRNKNKMKASKIKEIETISDSÖPÖSESDDEEDNDDELVSYN  
HLHKIKTEKENEPEDPIDIKKEAEKREKTRKEMKRLRKRKEKEDKELKEMK  
NKEKREKNVSESTÖKEKEVEKEKNAEMSEKKEKSIPLDITVHTTASSEPILSEK  
ESELHÖKTFESVCKERMPYKALÖKODRPEKGSSEEOLEHTROCILIGDITTECI  
KEYTNEÖIKÖRMKNILIFYKPEEPKARILHKLYKAEIKRÖSOÖHONÖISNVN  
THYTRNDVERLAKTETHHDSÖRSYSDSRLSYIHHDHNRÖGDAYKKSÖKRRPT  
SARSNSGHDHSDWHTYKODSRYISDSKRRKLDHRSRÖHRSRLEBNLDSRSHÖRSRSH  
SDHSHSDHRSSTSEYSHHKSRYDRYHSHDWÖMÖHRAÖSGSPRSLDÖRSPYGSPLGÖ  
HRSPFHSÖDHKSTPEHTWSSRKT"

misc\_feature  
/gene="CHD-2"  
/Note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1233 c 1520 g 1683 t  
ORIGIN

alignment\_scores:  
Quality: 197.00 Length: 41  
Ratio: 4.805 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 95.122

alignment\_block:  
US-08-973-363-9 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 ltleuProAspAspProAspLysLysProGlnAlaLysGlnleuGlnth 17  
|||||  
4080 ATTTTTACCTGATGATCCAGACAAGAAACCCGACGCAACAGCTACAGAC 4129  
34 luvAlaGlnArgLeuThrGlyAla 41  
|||||  
4130 CGGTGAGAGCTACTCATTAAATTACTGATGAATAAGACCTTGCAAGAAAG 4179  
17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLys 34  
|||||  
24180 AAGCACAAGAGCTGCTGCTGTGCA 4202

seq\_name: gb\_pat:A58686

seq\_documentation\_block:  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;

```

COMMENT      ISIS INNOVATION (GB)
FEATURES     Other publication AU 5906996 961224.
              Location/Qualifiers
              source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   60 a      34 c      31 g      28 t
ORIGIN
alignment_scores:
      Quality: 191.00      Length: 51
      Ratio: 4.659      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 80.392

Alignment_block:
US-08-973-363-9 x AS8686      ..

Align seg 1/1 to: AS8686 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasnLysAspLeuAlaArgLysgluValGlnArgLysLeuTrgly 40
|||||
101 AATTACTGATAATAAGACCTTGCAGAGAAAAAGAGTGCAGAAAGCCTTACTG 150
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:AS8684

seq_documentation_block:
LOCUS      AS8684      153 bp      DNA      linear      PAT 06-MAR-1996
DEFINITION AS8684 Sequence 3 from Patent WO9639505.
ACCESSION   AS8684
VERSION     AS8684.1 GI:3714247
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified.
REFERENCE   1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tjwari,B.
TITLE       AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL     Patent: WO 9639505-A 3 12-DEC-1996;
            ISIS INNOVATION (GB)
COMMENT     Other publication AU 5906996 961224.
FEATURES
            Location/Qualifiers
            source          1..153
                               /organism="unidentified"
                               /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
Percent Similarity: 80.392      Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x AS8684      ..

Align seg 1/1 to: AS8684 from: 1 to: 153

1 HleuProaspasProasp.....7
1 HleuProaspasProasp.....7
1 ATTTTACCTGATGACCCAGATTAAGAAACACAGGCAAGCATTTGCAGAC 50
8 LysLysProclinalalysglinleuglnthraTgalaasPTyrlleu124
|||||
51 CAAAGAACACAGGCAACAAACAGCATTTGCAGACCCGTGCAGATTAACCTCATTA 100
24 yslleuLeuasn
```

```

1 ATTTTACCTGATGATCCAGTAAAGCCAGGCAAGACGCTTACAGAC 50
|||||
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||
51 CAAAGAACCCAGGCAAGACGACTTACAGACCGCTGCAGACTTACTCATTA 100
|||||
24 LysLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuPheArgly 40
|||||
101 AATTACTGATTAAGACCTTGCAGAAAGGAGACCAAGGCTTGTCTGCT 150
|||||
41 Ala 41
|||
151 GCA 153

seq_name: gp_pat:A58685

seq_documentation_block:
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 4 from Patent WO9639505.
ACCESSION A58685
VERSION A58685.1 GI:3714248
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths R. and Thewari B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES
source other publication AU 5906996 961224.
location/Qualifiers
1..153
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t
ORIGIN

Alignment_scores:
Quality: 182.00 Length: 51
Ratio: 4.439 Gaps: 1
Percent Similarity: 80.392 Percent Identity: 76.471

Alignment_block:
US-08-973-363-9 x A58685, ..

Align seq 1/1 to: A58685 from: 1 to: 153

1 lleLeuProAspAspProAsp..... 7
|||||
1 ATTTTACCTGATGATCCAGTAAAGCCAGGCTTACAGACGACTTACAGAC 50
|||||
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
|||||
51 CAAAGAACCCAGGCTTACAGACGCTGCAGACTTACTCATTA 100
|||||
24 LysLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuPheArgly 40
|||||
101 AATTACTGATTAAGACCTTGCAGAAAGGAGACCAAGGCTTGTCTGCT 150
|||||
41 Ala 41
|||
151 GCA 153

seq_name: gp_to:MUSCHDIX

seq_documentation_block:
LOCUS MUSCHDIX 5349 bp mRNA linear ROD 19-SEP-1996
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.
ACCESSION L10410 X66028
```

VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 5349)  
TITLE Delmas,V., Stokes,D.G. and Perry,R.P.  
JOURNAL A mammalian DNA-binding protein that contains a chromodomain and an  
MEDLINE SNF2/SWI2-like helicase domain  
REFERENCE Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
AUTHORS 93211972  
TITLE 2 (bases 1 to 5349)  
JOURNAL Direct Submission  
Perry,R.P.  
COMMENT Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
FEATURES  
source Location/Qualifiers  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/translation="MNGHSDDEEVNNGSGSSQSDGDCGASGSGSSGSSSSDSS  
SOSGSDSDSGSGSSGSESDTSRENNVAKPKPKVDAEEMKSPSLIAVORSAML  
RKPOGAAQOORPASSNGSGSESDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS  
SDSESEERDKSSCCCTESDTPPKNVSRKQNSKSKNGKILIGQKRQIDSSDE  
DDEDVNDKRSRRQATVNVSYKDEDEMEKTDSDILEVCGEDVPQPEDEFEETIEM  
DCRVGRKATGATTIYAVEADGPNAGNERKDEGDIOYLIMKGMSHIHWTEEE  
TLKOQNVGRKMLDNYKKKQDETKRMKNASPEDEYVNCQOELTDDLKQYIARI  
IAHSNOKSAAGLPDYCKMOGLPYSECSMEDGALISKFPOTCIDEFESRNSKTPFK  
DCKVKORPRFALKKOPSYIGHGELEROLVINGIMLAHSNOKSGNSCIADENGL  
CKTIOTISLANTFHEHOLYGFLLVPLSTLTSNOREIQTASOMNAVYIGDINSR  
NMIRTHEMHPQTRKRWILITTYEILKDAFLGGLMWTGDEARLKNDSLL  
YKTLIDFNSNRLILTPPLQNSLAKELMSLHP IMEKRSSWDEEENKREGYA  
SLAKELEPLLRVKKVDEKSLPAVEQLLRNEMSLQOYKWLITRYKALISGSK  
GSTSGFLNIMELKKKCNHCYLIKPPDNEFYNKQELQHLISGKLLDLKLRL  
REGNRVLFSQWMLDILAYLYKROPPORLDSIGELGRKALDHPNNGSEDF  
CELLSTRAGGGINLASADTVIPDSMWPRNDLQAQAAHRTGCKQKNIRLYTKG  
SYEDILERAKKKMYLDILYVORMDTGKTIVLHSGSPSSSTPFNKEELATLKFEAE  
ELFKPEGEDEPOEMDIDELIKRAETHNEGPISVGDLLSOFKVANFSNMEDDI  
ELEPRNSKNMEIIPEDQRRLEEBEROKELLEYMLPRMNCAKOISFNGSEGRS  
RSRRYSGSDSDSISERKPKKGRPTIPRENIKGFSDAEIRRFISYKFGGPLERL  
DAIARDAELVDESETDLRLGLVHNGCYKALKDSSSGTERAGRLGKVKPTPLSG  
VOVNAKLYIAHDELILPKHSIPSDPEERKQYITICHTKAHPDIDMGKEDSNLLIG  
IYEGYGSHEMFKMPDLSLTKITLPDDPKRQAKOLQOTRADYILKLISPLAKREA  
QRLCGAGSKRRKTRAKSKAKMSIKVEEIKSDSPLSEKSDDEDDKLNDSPEK  
DRSKSVSDAVAHITASGEVPIAESEBELQKTFSLCKEEMRPYKALKQIDREK  
GISEREQLEHTROCLIKIGDHTTECLKEYNDEQIKQNMNIMIEVSKTEFDRKLH  
KLYHAIKKROESQONSQNSVATTHVIRNPMELKQNTMHNDSSRSYSDDRLLS  
OYHDHDKRHOGDSYKKSRSRKYSSPNSGNGKDHSMHRYRODSRYSPREKRLDD  
HRSRHRPSLDEGLDKDRCHSDHRSMDHMSDHRSSSHTHHKSRRDRLYSMDOLD  
HMASSGSPSPLDORSPYGSRPFESHAENRSTPEHTWSSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 182.00 Length: 41  
Ratio: 4.550 Gaps: 0  
Percent Similarity: 97.561 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-9 x MUSCHD1X ..

Align seg 1/1 to: MUSCHD1X from: 1 to: 5349

1 ILEuPProAspPProAspPlySProGlnAlaLysGlnLeuInTh 17  
|||||  
4026 ATGTTCTCCAGATGATCCTGATTAACCAACCAACCAACGTTACAGAC 4075  
17 FARGAlAspTYrLeuIleuLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4076 CCGGCGAGACTACTCATCAACACTACTTAGCAGAGATCTTGCAAAAAGAG 4125  
34 luvAlGlnArgLeuThGlyAla 41  
|||||  
4126 AGGCTCAGAGACTTGTGTGTCGCG 4148

seq\_name: gb\_pat:A58683

seq\_documentation\_block:  
LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths R. and Tjvari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES  
source Location/Qualifiers  
1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 167.00 Length: 51  
Ratio: 4.175 Gaps: 1  
Percent Similarity: 78.431 Percent Identity: 68.627

alignment\_block:  
US-08-973-363-9 x A58683 ..

Align seg 1/1 to: A58683 from: 1 to: 153

1 ILEuPProAspPProAspP.....  
1 ATTTCTCCAGATGATCCTGATTAACCAACCAACCAACGTTACAGAC 50  
8 LysLysProGlnAlaLysGlnLeuInThArgAlaAspTYrLeuIlel 24  
|||||  
51 CAAACCAACCAACCAACCAACGTTACAGACCCGTCAGACTACTCATCA 100  
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnLuvAlGlnArgLeuThGly 40  
|||||  
101 AACTACTTAGCAGAGATCTTGCAAAAAGAGGCTCAGAGACTTGTGTGT 150  
41 Ala 41  
|||  
151 GCG 153

seq\_name: gb\_pr:AF006513

seq\_documentation\_block:  
LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997  
DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428



[illegible]

```

1 |leleuProaspsPrroaspLysPrcogInalAtyScLnclnclnth 17  
| ||||||||  
4025 ATCTTCACGATGCATCCCATMAAAACCAACAAGCAAACGGTTCAGAC 4074  
  
17 ratGalAsPTyeUlleAltyLeauAnLynSglnLeuAlArLySG 34  
| ||||||||  
4075 CGGTGACAGACTACCCTCATTCAAAATTACTTGTAAGATCTTGCAAAAAAAG 4124  
  
34 luValGlmlArgIueThrglylNA 41  
| :::: ||::::||  
4125 AACCT.....CTTTCTGCTGG 4141  
  
seg_name: gb_pr:AC092372  
  
seq_documentation_block:  
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
ACCESSION AC092372  
VERSION AC092372.3 GI:17402768  
KEYWORDS htc.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 101220)  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
REFERENCE  
AUTHORS 2 (bases 1 to 101220)  
JOURNAL Unpublished  
TITLE DOE Joint Genome Institute.  
REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 101220)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission  
TITLE Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Pipaq Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146,7kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.  
Location/Qualifiers  
1..101220  
    /organism="Homo sapiens"  
    /db_xref="taxon:9606"  
    /chromosome="5"  
    /clone="RP11-58M12"  
BASE COUNT         34122 a   18862 c   17827 g   30409 t  
ORIGIN  
  
FEATURES  
source  
  
alignment_scores:  
    Quality:   166.00                  Length:   41  
    Ratio:     4.256                  Gaps:     1  
Percent Similarity: 95.122       Percent Identity: 82.927  
  
alignment_block:  
US-08-973-363-9 x AC092372/rev ..  
  
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220  
  
1 lleleuProaspsPrroaspLysPrcogInalAtyScLnclnclnth 17  
| ||||||||  
26932 ATTCTTCAGATGCATCCCATMAAAACCAACAAGCAAACGGTTCAGAC 26983
```

```

17   ratgalaasptyrleuilelyseuleuasnlysaspleualaargysg 34
|||||.....|
26882 CCCTGCAGACTACCTCATCAATTAATTACTTAGAGATCTTGCAAAAAAAG 26833
34   luvalgalnargleuthnrglyala 41
||:::|||||
||:::|||||
26832 AAGCT.....CTTCTGCGTCG 26816

seq_name: gb_pr:AC012624

seq_documentation_block:
LOCUS      AC012624                134365 bp    DNA          linear   PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION  AC012624
VERSION    AC012624.6   GI:14993679
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 134365)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 134365)
AUTHORS   DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL    Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3  (bases 1 to 134365)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            4  (bases 1 to 134365)
AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL    Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Jul 21, 2001 this sequence version replaced gi:14277267.

COMMENT    Location/Qualifiers
FEATURES   1..134365
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN
alignment_scores:
Quality: 166.00 Length: 41
Ratio: 4.256 Gaps: 1
Percent Simlarity: 95.122 Percent Identity: 82.927

alignment_block:
US-08-973-363-9 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365
1 11eleuprohaspproaspplyslysproglalalysglinleuglnrh 17
118247 ATTCCTTCGAGATGATCCCGATATAAAAACCAAGCAAAGAAGTTGCCGAC 118296
17 ratgalaasptyrleuilelyseuleuasnlysaspleualaarglsygc 34
|||||.....|
118297 CCCTGCAGACTACCTCATCAAAATTACTTAAGTAGAGATCTTGCAAAAAAAG 118346
34 luvalgalnargleuthnrglyala 41
||:::|||||
||:::|||||
118347 AAGCT.....CTTCTGCGTCG 118363

seq_name: gb_htcg:AC021449
```

```

LOCUS       documentacion.block:                               143079 bp    DNA        linear   HTG-10-SEP-2000
DEFINITION  Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
ACCESSION   ACO21449
VERSION     ACO21449.3  GI:10047806
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143079)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL     Unpublished
TITLE       2  (bases 1 to 143079)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckert,L.R., Beda,F.,
            Bonuslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Chaoel,y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Dekarrellano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McKean,P., McGurt,A., McKernan,K.,
            McPheters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Plierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Titrrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,M.J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Sep 10, 2000 this sequence version replaced gi:7407963.
            All repeats were identified using RepeatMasker:
            Smit, A.P.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L5154
            Center clone name: 58_M12
            ----- Summary Statistics
            Sequencing vector: M13; M7815; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 134743 bases at least Q40
            Consensus quality: 139227 bases at least Q30
            Consensus quality: 140814 bases at least Q20
            Insert size: 14400; agarose-fp
            Insert size: 142179; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 38820: contig of 38820 bp in length
            * 38821 38920: gap of 100 bp
            * 38921 40411: contig of 1491 bp in length
            * 40412 40511: gap of 100 bp

```



Ratio: 4.256 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 82.927  
alignment block:  
US-08-973-363-9 x AC008531 ..  
Align seg 1/1 to: AC008531 from: 1 to: 145659  
1 lleuProaspProaspLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
46770 ATTCTTCAGATGATCCGATAAAAACACACAAACAGCTGCAGAC 46819  
17 rArgAlaSpTyrLeuIleuLysLeuAaSnLysAspLeuAlaArgLysG 34  
|||||  
46820 CCGTCAGACACTCCCTCATCAATTACTAGAGAGATCTTCAAAAAAG 46869  
34 luValGlnArgLeuThrGlyAla 41  
|||||  
46870 AAGCT.....CTTCTGTGTCG 46886  
seq\_name: gb\_htg:AC091946  
seq\_documentation block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 193446)  
TITLE DOE Joint Genome Institute.  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 193446)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
AUTHORS Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCI-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-1p estimation  
Estimated insert size: 190246; sum-of-contigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-1p estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3823 3923: gap of unknown length  
\* 3924

5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12560: contig of 2100 bp in length  
\* 12561 14611: gap of unknown length  
\* 14612 14711: contig of 2051 bp in length  
\* 14712 16381: gap of unknown length  
\* 16382 16481: contig of 1670 bp in length  
\* 16482 17968: gap of unknown length  
\* 17969 18068: contig of 1487 bp in length  
\* 18069 20434: gap of unknown length  
\* 20435 20534: contig of 2366 bp in length  
\* 20535 23515: gap of unknown length  
\* 23516 23615: contig of 2981 bp in length  
\* 23616 27563: gap of unknown length  
\* 27564 30987: contig of 3948 bp in length  
\* 30988 31087: gap of unknown length  
\* 31088 36065: contig of 3324 bp in length  
\* 36066 36165: gap of unknown length  
\* 36166 40978: gap of unknown length  
\* 40979 41078: contig of 4813 bp in length  
\* 41079 45663: gap of unknown length  
\* 45664 45763: contig of 4585 bp in length  
\* 45764 51745: gap of unknown length  
\* 51746 51845: contig of 5982 bp in length  
\* 51846 57359: gap of unknown length  
\* 57360 57459: contig of 5514 bp in length  
\* 57460 67881: gap of unknown length  
\* 67882 67981: contig of 10422 bp in length  
\* 67982 74132: gap of unknown length  
\* 74133 74232: contig of 6151 bp in length  
\* 74233 79795: gap of unknown length  
\* 79796 79895: contig of 5563 bp in length  
\* 79896 87511: gap of unknown length  
\* 87512 87611: contig of 7616 bp in length  
\* 87612 92791: gap of unknown length  
\* 92792 92891: contig of 5180 bp in length  
\* 92892 102794: gap of unknown length  
\* 102795 102894: contig of 9903 bp in length  
\* 102895 110866: gap of unknown length  
\* 110867 110966: contig of 7972 bp in length  
\* 110967 117571: gap of unknown length  
\* 117572 117671: contig of 6605 bp in length  
\* 117672 123738: gap of unknown length  
\* 123739 123838: contig of 6067 bp in length  
\* 123839 130583: gap of unknown length  
\* 130584 130683: contig of 6745 bp in length  
\* 130684 141544: gap of unknown length  
\* 141545 141644: contig of 10861 bp in length  
\* 141645 169109: gap of unknown length  
\* 169110 169209: contig of 27465 bp in length  
\* 169210 193446: gap of unknown length  
\* 193446 24237: contig of 24237 bp in length.  
Location/Qualifiers  
1..193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_11b="RPCI Human BAC library 11"

BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN

alignment\_scores: 166.00 Length: 41  
Quality: 4.256 Gaps: 1  
Ratio: 4.256



---



34 luValGlnArgLeuThrGlyAla 41  
 1238 AACGACAGAGACTTGTCTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

AA142751;

12-MAR-1997 (first entry)

Chicken CHD-1A gene.

Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

Gallus sp.

Key Location/Qualifiers

FT CDS 228..5390

FT /tag= a

PN WO9639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Claim 1; Fig 5; 76pp; English.

The chicken CHD-W gene (AA142754) acting alone or in conjunction with

the closely related CHD-1A gene (AA142751) is suggested to initiate

female development in birds. The sequence of CHD-1A was deduced

from 3 clones isolated from a stage 10-12 chicken embryo cDNA

library using a great tit CHD-W sequence (see also AA142755) as probe.

The CHD-1A (A - Avian) gene shows close identity to the mouse CHD-1

gene (see also AA142756-57). It is located on an autosome or Z

chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

specific signal on hybridisation to genomic DNA of a non-ratite

bird and can be used for sex determin. of a bird. CHD-1A nucleic

acids can also be used to control the sex of the progeny of a bird.

Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores: Quality: 197.00 Length: 41

Percent Similarity: 100.000 Percent Identity: 95.122

US-08-973-363-9 x AA142751

Align seg 1/1 to: AA142751 from: 1 to: 6608

|||||  
 4130 CCGTGCAGACTACCTCAATTAATTAAGAAAGCAAGG 4179  
 34 luValGlnArgLeuThrGlyAla 41  
 4180 AAGCACAAGAGCTTGTCTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq\_documentation\_block:

ID AA142759 standard; DNA; 153 BP.

AA142759;

12-MAR-1997 (first entry)

Great tit CHD-W gene fragment.

Bird; sex determination; chromodomain-Helicase-DNA binding 1;

CHD-1A; CHD-W; W chromosome; ss.

Parus major.

Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"

PN WO9639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Claim 8; Fig 3; 76pp; English.

Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

to portions of the chicken CHD-1A (A - Avian) gene (AA142757),

CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and

CC and the great tit CHD-W gene (AA142759). Translated amino acid

sequences of this region are provided in AA08146-49. The CHD-1A

(see also AA142751) and CHD-W (see also AA142754-55) genes determine

sex in birds and can be used to identify the sex of an embryo,

foetus etc. and to manipulate the sex of progeny.

Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment\_scores: Quality: 191.00 Length: 51

Percent Similarity: 80.392 Percent Identity: 80.392

US-08-973-363-9 x AA142759

Align seg 1/1 to: AA142759 from: 1 to: 153

1 lleuProaspPaspProaspLysLysProGlnAlaLysGlnLeuGlnrh 17



```

|||||
1 ATTTACTGATGATTCAGATTAAGAAACCAAGCAAGCACTTGCAGAC 50
8 .LysLeuProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlel 24
|||||
51 CAAGAAACCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAATTA 100
24 yLsLeuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuThrGly 40
|||||
101 AATTACTGATTAAGACCTTGCAAGAAAGAAAGAAAGCAAGCAAGCAAGCTTGT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757
seq_documentation_block:
ID AA142757 standard; DNA: 153 BP.
AC AA142757;
XX
XX 12-MAR-1997 (first entry)
DE Chick CHD-1A gene fragment.
XX
XX Bird: sex determination: chromodomain-Helicase-DNA binding 1;
KM CHD-1A; CHD-W; W chromosome; ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH msc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
PI
XX WPI: 1997-043127/04.
DR P-PSDB: AAM08147.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757).
CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)
CC and the great tit CHD-W gene (see also AA142759). Translated amino
CC acid sequences of this region are provided in AAM08146-49. The
CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes
CC determine sex in birds and can be used to identify the sex of an
CC embryo, foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment_scores:
Quality: 182.00 Length: 51
Ratio: 4.439 Gaps: 1

Percent Similarity: 80.392 Percent Identity: 76.471
Alignment_block:
US-08-973-363-9 x AA142757
Align seq 1/1 to: AA142757 from: 1 to: 153

1 TleLeuProAspPaspProAsp..... 7
|||||
1 ATTTACTGATGATTCAGATTCAGAAACCCAGCAAGCAAGCAAGCAAGCAAGCA 50
8 .LysLeuProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlel 24
|||||
51 CAAGAAACCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAATTA 100
24 yLsLeuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuThrGly 40
|||||
101 AATTACTGATTAAGACCTTGCAAGAAAGAAAGCAAGCAAGCAAGCAAGCTTGT 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA: 153 BP.
AC AA142758;
XX
XX 12-MAR-1997 (first entry)
DE Chick CHD-W gene fragment.
XX
XX Bird: sex determination: chromodomain-Helicase-DNA binding 1;
KM CHD-1A; CHD-W; W chromosome; ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH msc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
PI
XX WPI: 1997-043127/04.
DR P-PSDB: AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757).
CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC and the great tit CHD-W gene (AA142759). Translated amino acid
CC sequences of this region are provided in AAM08146-49. The CHD-1A
CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,

```

CC foetus etc. and to manipulate the sex of progeny.  
XX  
SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;  
  
alignment\_scores:  
Quality: 182.00 Length: 51  
Ratio: 4.439 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 76.471  
  
alignment\_block:  
US-08-973-363-9 x AAT42758 ..  
  
Align seg 1/1 to: AAT42758 from: 1 to: 153  
  
1 lleuProaspaspProasp..... 7  
1 ATTTCCTCATGATGATCCAGATGAAGAAACCCAGCTAAGCAGTTACAGAC 50  
  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
51 CAAAGAAACCCAGCGCTAAGCAGTTACAGACCCGTCAGATTCATCTTA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGluValGlnArgLeuThGly 40  
101 AATTACTGAATAAAGACCTTGCAAGAAAGCAGACGACTTGCTGT 150  
41 Ala 41  
151 GCA 153  
  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756  
seq\_documentation\_block:  
ID AAT42756 standard; DNA; 153 BP.  
XX  
AC AAT42756;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Mouse CHD-1 gene (bases 3855-9777).  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KM CHD-1; CHD-W; W chromosome; ss.  
XX  
OS Mus sp.  
XX  
FH Key location/Qualifiers  
FT misc\_difference 52..81  
FT /\*tag= a  
FT and are ignored in the translated amino  
FT acid sequence given in Fig 3"  
XX  
PN W09639505-A1.  
XX  
PD 12-DEC-1996.  
XX  
PF 05-JUN-1996; 96WO-GB01341.  
XX  
PR 06-JUN-1995; 95GB-0011439.  
XX  
PA (ISIS-) ISIS INNOVATION LTD.  
XX  
PI Griffiths R, Tiwari B;  
XX  
DR WPI: 1997-043127/04.  
DR P-PSDB; AAW08146.  
XX  
PT Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determ. and to control sex of progeny  
XX  
PS Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
CC and AAT42757) chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of  
CC progeny.  
XX  
SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
  
alignment\_scores:  
Quality: 167.00 Length: 51  
Ratio: 4.175 Gaps: 1  
Percent Similarity: 78.431 Percent Identity: 68.627  
  
alignment\_block:  
US-08-973-363-9 x AAT42756 ..  
  
Align seg 1/1 to: AAT42756 from: 1 to: 153  
  
1 lleuProaspaspProasp..... 7  
1 ATTCTCCAGATGATCCTGATTAACCAACCAACCAAGCAGTTACAGAC 50  
  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
51 CAAAGAAACCCAGCGCTAAGCAGTTACAGACCCGTCAGACTTACCTCATCA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGluValGlnArgLeuThGly 40  
101 AACTACTTACGACGATCTTCACAAAGAGAGCGCTCAGAGACTTGTGGT 150  
41 Ala 41  
151 GCG 153  
  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280  
seq\_documentation\_block:  
ID AAV59280 standard; cDNA; 1311 BP.  
XX  
AC AAV59280;  
XX  
DT 14-DEC-1998 (first entry)  
XX  
DE Altered telomere repeat binding factor 1 gene.  
XX  
KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
KM telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
XX  
OS Homo sapiens.  
XX  
OS Synthetic.  
XX  
FH Key location/Qualifiers  
FT CDS 1..1311  
FT /\*tag= a  
FT /product= "A-TRF"  
XX  
PN W09836066-A1.  
XX  
PD 20-AUG-1998.  
XX  
PF 13-FEB-1998; 98WO-US02765.  
XX  
PR 04-FEB-1998; 98US-0018628.  
PR 13-FEB-1997; 97US-0800264.  
XX  
PA (UYRO ) UNIV ROCKEFELLER.  
XX

PI Bianchi A, De Lange T, Van Steensel B;  
 XX MPI: 1998-480769/41.  
 DR P-ESDB: AAV59280.  
 XX  
 PT Nucleic acid encoding altered telomere repeat binding protein and  
 related vectors - transfectants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 XX  
 PS Claim 14: Page 110-111; 163pp; English.  
 XX  
 CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
 CC inhibit shortening of telomeres associated with ageing (for cosmetic  
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
 CC tumours and viral (including human immunodeficiency virus) infection.  
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
 CC for expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.  
 CC  
 SO Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;  
 XX  
 alignment\_scores:  
 Quality: 129.00 Length: 25  
 Ratio: 5.160 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-9 x AAV59280 ..  
 Align seg 1/1 to: AAV59280 from: 1 to: 1311  
 1 TleleuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnth 17  
 |||||||  
 1237 ATTCTTCAGATGATCTCATATAAACCAAGCAAGCAAGCAAGTACAGAC 1286  
 |||||||  
 17 rArGaLaAspTyrLeuIleLysLeu 25  
 |||||||  
 1287 CCGTCACAGCTACCTCATCTCAACTTA 1311  
 seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882  
 seq\_documentation\_block:  
 ID AAK88882 standard; CDNA; 421 BP.  
 XX  
 AC AAK88882;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
 XX  
 KM Human: digestive system antigen; gene therapy; cancer; appendicitis;  
 KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KM digestive system disorder; Meckel's diverticulum; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155314-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 17-JAN-2001; 2001WO-US01324.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-MAR-2000; 2000US-0190122.  
 PR 19-MAR-2000; 2000US-0205313.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214866.  
 PR 30-JUN-2000; 2000US-0215133.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0228287.  
 PR 01-SEP-2000; 2000US-0228343.  
 PR 01-SEP-2000; 2000US-0228344.  
 PR 01-SEP-2000; 2000US-0228345.  
 PR 05-SEP-2000; 2000US-0228509.  
 PR 05-SEP-2000; 2000US-0228513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235488.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 28-SEP-2000; 2000US-0236327.  
 PR 28-SEP-2000; 2000US-0236367.  
 PR 28-SEP-2000; 2000US-0236368.  
 PR 28-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244747.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI: 2001-502630/55.  
DR P-PSDB; AAM93109.  
XX  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosis, treating, preventing and/or prognostic disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Claim 1; SEQ ID NO 1198; 986bp; English.  
XX  
CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:  
Quality: 120.00 Length: 33  
Ratio: 4.138 Gaps: 0  
Percent Similarity: 87.879 Percent Identity: 72.727  
  
alignment\_block:  
US-08-973-363-9 x AAK88882 ..  
  
Align seq 1/1 to: AAK88882 from: 1 to: 421  
  
1 HleuPProAspProAspPlyLysProGlnAlaLysGlnLeuGlnH 17  
|||||||  
110 APTCTGCGCGTGGACACAGATAAAAGCCTCAGGGAGACGACTACGAC 159  
  
17 FARGAlAsPTyRleuLleuLysLeuLeuAsnLysAsPLeuAlaArgLys 33  
|||||||  
160 CCGAGCGGATTACTTGTGTAAGCTGCTCAGAAAGCTGTGAGAGAGAG 208  
  
seq\_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI57603  
  
seq\_documentation\_block:  
ID AAI57603 standard; cDNA; 421 BP.  
XX  
AC AAI57603;  
XX  
DT 19-OCT-2001 (first entry)  
XX  
DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.  
XX  
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155350-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US01350.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 16-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.

PR	14-NOV-2000	2000US	-0.2252678	PR	14-NOV-2000	2000US	-0.2252678
PR	14-NOV-2000	2000US	-0.2252547	PR	14-NOV-2000	2000US	-0.2252547
PR	14-NOV-2000	2000US	-0.2252575	PR	14-NOV-2000	2000US	-0.2252575
PR	14-NOV-2000	2000US	-0.2252578	PR	14-NOV-2000	2000US	-0.2252578
PR	18-NOV-2000	2000US	-0.2256279	PR	18-NOV-2000	2000US	-0.2256279
PR	22-NOV-2000	2000US	-0.2256671	PR	22-NOV-2000	2000US	-0.2256671
PR	22-NOV-2000	2000US	-0.2256868	PR	22-NOV-2000	2000US	-0.2256868
PR	23-AUG-2000	2000US	-0.2271702	PR	23-AUG-2000	2000US	-0.2271702
PR	30-AUG-2000	2000US	-0.2282827	PR	30-AUG-2000	2000US	-0.2282827
PR	01-SEP-2000	2000US	-0.2292934	PR	01-SEP-2000	2000US	-0.2292934
PR	01-SEP-2000	2000US	-0.2293344	PR	01-SEP-2000	2000US	-0.2293344
PR	01-SEP-2000	2000US	-0.2293509	PR	01-SEP-2000	2000US	-0.2293509
PR	05-SEP-2000	2000US	-0.2293413	PR	05-SEP-2000	2000US	-0.2293413
PR	06-SEP-2000	2000US	-0.2304348	PR	06-SEP-2000	2000US	-0.2304348
PR	08-SEP-2000	2000US	-0.2312442	PR	08-SEP-2000	2000US	-0.2312442
PR	08-SEP-2000	2000US	-0.2312444	PR	08-SEP-2000	2000US	-0.2312444
PR	08-SEP-2000	2000US	-0.2314143	PR	08-SEP-2000	2000US	-0.2314143
PR	08-SEP-2000	2000US	-0.2314144	PR	08-SEP-2000	2000US	-0.2314144
PR	08-SEP-2000	2000US	-0.2320801	PR	08-SEP-2000	2000US	-0.2320801
PR	12-SEP-2000	2000US	-0.2321967	PR	12-SEP-2000	2000US	-0.2321967
PR	14-SEP-2000	2000US	-0.2323297	PR	14-SEP-2000	2000US	-0.2323297
PR	14-SEP-2000	2000US	-0.2323298	PR	14-SEP-2000	2000US	-0.2323298
PR	14-SEP-2000	2000US	-0.2323401	PR	14-SEP-2000	2000US	-0.2323401
PR	14-SEP-2000	2000US	-0.2330663	PR	14-SEP-2000	2000US	-0.2330663
PR	14-SEP-2000	2000US	-0.2330673	PR	14-SEP-2000	2000US	-0.2330673
PR	21-SEP-2000	2000US	-0.2342323	PR	21-SEP-2000	2000US	-0.2342323
PR	21-SEP-2000	2000US	-0.2342474	PR	21-SEP-2000	2000US	-0.2342474
PR	23-SEP-2000	2000US	-0.2344597	PR	23-SEP-2000	2000US	-0.2344597
PR	26-SEP-2000	2000US	-0.2345884	PR	26-SEP-2000	2000US	-0.2345884
PR	27-SEP-2000	2000US	-0.2325384	PR	27-SEP-2000	2000US	-0.2325384
PR	29-SEP-2000	2000US	-0.2326327	PR	29-SEP-2000	2000US	-0.2326327
PR	29-SEP-2000	2000US	-0.2326357	PR	29-SEP-2000	2000US	-0.2326357
PR	29-SEP-2000	2000US	-0.2326368	PR	29-SEP-2000	2000US	-0.2326368
PR	29-SEP-2000	2000US	-0.2326370	PR	29-SEP-2000	2000US	-0.2326370
PR	02-OCT-2000	2000US	-0.2326802	PR	02-OCT-2000	2000US	-0.2326802
PR	02-OCT-2000	2000US	-0.2327038	PR	02-OCT-2000	2000US	-0.2327038
PR	02-OCT-2000	2000US	-0.2327038	PR	02-OCT-2000	2000US	-0.2327038
PR	02-OCT-2000	2000US	-0.2327038	PR	02-OCT-2000	2000US	-0.2327038
PR	13-OCT-2000	2000US	-0.2327935	PR	13-OCT-2000	2000US	-0.2327935
PR	13-OCT-2000	2000US	-0.2329337	PR	13-OCT-2000	2000US	-0.2329337
PR	20-OCT-2000	2000US	-0.2329630	PR	20-OCT-2000	2000US	-0.2329630
PR	20-OCT-2000	2000US	-0.2411281	PR	20-OCT-2000	2000US	-0.2411281
PR	20-OCT-2000	2000US	-0.2411785	PR	20-OCT-2000	2000US	-0.2411785
PR	20-OCT-2000	2000US	-0.2417187	PR	20-OCT-2000	2000US	-0.2417187
PR	20-OCT-2000	2000US					

[illegible]

(HUMA-) HUMAN GENOME SCI INC

Rosen CA, Barash SC, Ruben SM,

WPI: 2001-457727/49.

P-PSDB; AAM38625.

Isolated polypeptide for treating, preventing and/ or prognosing disorders related to the colon and rectum including colorectal cancers and also for testing and detection e.g. diagnosis -

Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English

The present invention provides the protein and coding sequences of a number of colorectal cancer antigens. These are shown in **AA157547-AA157619** and **AA83569-AA838641**. These can be used in the diagnosis, prevention and treatment of cancer of the colon and/or rectum. The sequence is a colorectal cancer antigen coding sequence of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/publications-pct-sequences](http://wipo.int/pub/publications-pct-sequences).

Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other,

alignment_scores:	Length:
Quality:	120.00
Ratio:	4.138
Percent Similarity:	87.879
Percent Identity:	72.722
alignment_block:	
US-08-973-363-9 x AA157603	..

Align seg 1/1 to: AAI57603 from: 1 to: 421

```

1 1lleuPProaspPProaspLysProGlnAlaLysGlnLeuGlnTh 17
||||| ..... |||||||||..... |||||||||
110 ATTCTGCCGCTGGACACATATAAAGCCTCAGGGAGCAGCTACAGAC 159
17 rArgAlaSPtyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLys 33
|||||..... ||| ..... |||
160 CCGAGCGGATTTACTTGTGAAGCTGCTCAGAAAGGCTGTGAGAGAGAAG 208

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:

ID ABL06443 standard; cDNA; 6240 BP.

AC ABL06443;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 13811.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR P-PSDB; ABB62340.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Claim 1: SEQ ID NO 13811; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (AB057737-AB072072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

alignment\_scores:

Quality: 87.50 Length: 30  
 Ratio: 3.241 Gaps: 1  
 Percent Similarity: 90.000 Percent Identity: 60.000

alignment\_block:

US-08-973-363-9 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```

1 1lleuPProaspPProaspLysProGlnAlaLysGlnLeuGlnTh 17
||||| ..... |||||||||..... |||||||||
4184 ATTTCGCTTAACGAT...ACGCGCAAGCCCCAGGCCAAGCAGCTGCAGAC 4230

```

```

17 rArgAlaSPtyrLeuIleLysLeuLeuAsnLysAspLeu 30
|||||..... ||| ..... |||
4231 GCGTCCGAGTACCTGCTCAAGATCATCAAGAAAGACGTG 4270

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:

ID ABL06442 standard; cDNA; 9933 BP.

AC ABL06442;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 13808.

KW Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR P-PSDB; ABB62339.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Claim 1: SEQ ID NO 13808; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB057737-AB072072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

alignment\_scores:

Quality: 87.50 Length: 30  
 Ratio: 3.241 Gaps: 1  
 Percent Similarity: 90.000 Percent Identity: 60.000

alignment\_block:

US-08-973-363-9 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

1 1lleuPProaspPProaspLysProGlnAlaLysGlnLeuGlnTh 17







alignment\_scores:

Quality:	60.50	Length:	37
Ratio:	2.161	Caps:	1
Percent Similarity:	75.676	Percent Identity:	29.730

alignment\_block:

US-08-973-363-9 x ABI99578 ..

Align seg 1/1 to: ABI99578 from: 1 to: 826

```

2 LeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThrAf 18
|||||.....|.....|.....|.....|.....|.....|.....|
278 CTTCGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 327
18 gAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaargLysGlnV 35
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :
328 ATTTGAGAACCTCTGCAAAATTATG...AAGGATATTGTTGGAGAGAGAG 374
35 aLglnArgLeu 38
|||.....|.....|.....|.....|.....|.....|.....|
375 TTGAGAAAGGTC 385

```

**THIS PAGE BLANK (USP 11)**



```

seq_documentation_block:
LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001
DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone L1E1d12 5', mRNA
sequence.
ACCESSION AL644594
VERSION AL644594.1 GI:16796719
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE
1 (bases 1 to 645)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: L1E1d12.plc
Sequencing primer: PLC
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
source
1..645
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="L1E1d12"
/clone_lib="XGC-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT 222 a 125 c 156 g 141 t 1 others
ORIGIN
alignment_scores:
Quality: 183.00 Length: 39
Ratio: 4.816 Gaps: 0
Percent Similarity: 97.436 Percent Identity: 92.308
alignment_block:
US-08-973-363-9 x AL644594 ..
Align seg 1/1 to: AL644594 from: 1 to: 645
1 TTTCTTAAACGATGATCCAGATTAAGAGCCCAAGCCACAGCTACAGAC 17
|||||
456 ATTTTAAACGATGATCCAGATTAAGAGCCCAAGCCACAGCTACAGAC 505
|||||
17 TATGATAAAPTATTTTCTTAACTTCTCATTAAGATCTCTTACGAAAG 34
|||||
506 CAGAGCTGATCTTAACTTCTCATTAAGATCTCTTACGAAAG 555
|||||
34 TTTCTTAAACGATGATCCAGATTAAGAGCCCAAGCCACAGCTACAGAC 39
|||||
556 AAGACCAAGACTTCTTAACTTCTCATTAAGATCTCTTACGAAAG 572
|||||
seq_name: gb_est1.BB155356
seq_documentation_block:
LOCUS BB155356 619 bp mRNA linear EST 18-OCT-2001
DEFINITION BB155356 RIKEN full-length enriched, 16 days neonate thymus Mus
musculus cDNA clone A130024L16 3' similar to L10410 Mouse
DNA-binding protein (CHD-1) mRNA, mRNA sequence.
ACCESSION BB155356
VERSION BB155356.2 GI:16268254
KEYWORDS EST.

```

SOURCE	mouse mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 619)
AUTHORS	Atkawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Himoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakihara,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Atkawa,T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	On Jun 29, 2000 this sequence version replaced gi:1811286. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanahki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamakoshi,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Genomic Sciences Project of Genome Exploration Research Group in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Location/Qualifiers 1. 619 /organism="Mus musculus" /db_xref="taxon:10090" /clone="AI30024L16" /clone_id="RIKEN full-length enriched, 16 days neonate thymus" /tissue_type="thymus" /dev_stage="16 days neonate" /lab_host="DH10B" /note="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGAGAGATCCAGAGCCTTTTCTTTTCTTTTNN 3', cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAAATTAATATCCGCCGCCGCC3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBlueScript KS(+) after bulk excision from Lambda

```

BASF COUNT      194 a      118 c      161 g      146 t
ORIGIN
Alignment_scores:
      Quality: 182.00      Length: 41
      Ratio: 4.350      Gaps: 0
      Percent Similarity: 97.361      Percent Identity: 85.366
Alignment_block:
US-08-973-363-9 x BB155356 ..
Align seg 1/1 to: BB155356 from: 1 to: 619
1 11e1eProAaPaSPProAaSPlysLysProGlnAlaLysGlnLeuGln1nTh 17
|||||
489 ATTCTTCGAGATGATCTGATGATAAAAACACACAGCAAAACAGTTACAGAC 538
17 TargAlaAPtyrLeu11eLysLeuLeuLysAspLeuAlaArgLys 34
|||||
539 CCGTCAGACACTACCTCATCAACACTTACGACAGACTCTGCAAAAAGAG 588
34 LuvaGlnArgLeuThrGlyAla 41
|||||
589 AGGCTCAGAGACTTGTGTGTGG 611
seq_name: gb_est1:BB461065
seg_documentation_block:
LOCUS      BB461065      660 bp      mRNA      linear      EST 25-OCT-2001
DEFINITION      BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion
MUS musculus cDNA clone D13070B13.3 similar to L10410 Mouse
DNA cloning product (cDNA) mRNA, mRNA sequence.
ACCESSION      BB461065
VERSION        BB461065.2 GI:16426612
KEYWORDS
SOURCE        EST.
ORGANISM      house mouse.
MUS musculus
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 660)
AtekhAa,T., Carrincci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Kouda
K., Hiramoto,K., Hoti,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,H.,
D., Shibata,K., Shinagawa,A., Shitaki,T., Sogabe,Y., Suzuki,H.,
Tegami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arkawa,T., et al. 2001)
Unpublished (2001)
On Jul 21, 2000 this sequence version replaced gi:936558.
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shuhri-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-rsg@sc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Carrincci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
K., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Gene Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E.,
```

```

Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,I., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,X. and Hayashizaki,Y.  

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 261-269 (2001)  

Kondo,S., Shinagawa,A., Salto,T., Kiyosawa,R., Yamahata,I., Alzawawi,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  

Location/Qualifiers  

1..660  

/oranism="Mus musculus"  

/db_xref="taxon:10090"  

/clone_id="D13007Bj3"  

/clone_lib="RIKEN full-length enriched, 12 days embryo spinal ganglion"  

/tissue_type="spinal ganglion"  

/dev_stage="12 days embryo"  

/lab_host="DH10B"  

/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'  

GAGGAGAACAAGCGCCGCACACTCGAGATTTTGTCTTTTTVNN 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'  

GAGGAGAACAATTCTGGATTAAATAAATTATTCATCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified phluescript KS(+) after bulk excision from lambda FLC I."
BASE COUNT      211 a       130 c       168 g       151 t
ORIGIN
alignment_scores:  

    Quality:   182.00     Length:        41  

    Ratio:     4.550      Gaps:          0  

Percent Similarity: 97.561 Percent Identity: 85.366
Alignment_block:  

US-08-973-363-9 x BB461065 ..  

Align seg 1/1 to: BB461065 from: 1 to: 660  

1 TleueProaspsaprobaspulsylsgproginAlalsglnlenclnth 17  

|||||  

481 ATTCTTCAGTGATGCTCSTRTAAAACCAACAGGAAGAGTTTCAACAC 530  

|||||  

17 TAGTAAsptYrLeuIlElysleuAuasnlylsAspleualatarglySG 34  

|||||  

531 CGGTGACAGACTACCCTCATCAAACACTCTTAGCAGAGATCTTGCAAAAAAG 580  

|||||  

34 luvAlgnAtArgLeuthrlglala 41  

||:::||::|||  

581 AGGCTCACAGACTTGTGTGTGCTGCG 603

```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS	BB834922	446 bp	mRNA	linear	EST 19-NOV-2001
DEFINITION	BB834922	RIKEN full-length enriched, mammary gland RCB-0527			

Jyg-MC(B) CDNA Mus musculus CDNA clone G930033J21 3', mRNA sequence.

ACCESSION	BB834922
VERSION	BB834922.1
GI	GI:17013165

KEYWORDS	EST.
SOURCE	house mouse.

ORGANISM      Mus musculus  
Eukaryota; Metazoa; Chordata; Mammalia; Rodentia; Muridae; Mus

Mammalia; Eutheria; Rodentia	
1 (bases 1 to 446)	
<b>REFERENCE</b>	
Almouzni G, et al. (1987) <i>J Mol Biol</i> 194: 1-14.	

**AUTHORS**  
Akimura, T., Arakawa, T.,  
Hayatsu, N., Hiramoto, K.

Y., Ito, M., Kawai, J.,  
Nakamura, M., Nishi, K.,

Saito, R., Sakai, C., Shibata, K., Shinagawa, A.

A., Takahashi, F., Takahashi, A., Watahiki, A., Yasunishiki, A., and Takahashi, F.

TITLE	YEAR
RIKEN Encyclopedia of Materials Science and Engineering	2001)

JOURNAL COMMENT	unpublished (2001)
Contact: Yoshinide Haya	

Laboratory for Genome  
Sciences Center(GSC), Y

The Institute of Physics  
1-7-22 Suehiro-cho, Tsurumi  
Tel. 01-45-503-0000

Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: [genome-res@qsc.i](mailto:genome-res@qsc.i)  
URL: <http://genome.qsc.i>  
Contact: D. Chibata

Carninci, P., Shibata, Y.,  
, M., Konno, H., Okazaki,  
Normalization and sub

Normalization and sub  
prepare full-length cDN

genes. *Genome Res.* 10: 1001-1010.

Wataniki, M., Yoneda, Y.,  
S., Kawai, J., Okazaki,  
Havashizaki Y

RIKEN integrated sequencing pipeline with

sequencing pipeline with  
10 (11), 1757-1771 (2006)

Konho, H., Fukunishi, Y.,  
Y. and Hayashizaki, Y.  
Computer-based methods

computer-based methods  
encyclopedia: real-time  
nonredundant cDNA library

Please visit our web site for further details.

location/quality of mouse tissues.

```

1. .446
/organism="Mus
source
removed

```

```
/organism= Mus
/db_xref= "taxc
/ci= "G93003
```

```
/clone= 653003
/clone_lib="RI
RCB-0527 Jya-M
```

```
incd 0027 009 1
/tissue_type="
/cell_line="RC
```

BASE COUNT	139 a	99 c	1
ORIGIN			

FEATURES	source	Location/Qualifiers
		1..438
		/organism="Mus musculus"
		/db_xref="taxon:10090"
		/clone="G930013K04"
		/clone_lib="Riken full-length enriched, mammary gland
		RCB-0527 Jy9-MC(B) cDNA"
		/tissue_type="mammary gland"
		/cell_line="RCB-0527 Jy9-MC(B)"
		/cell_line="RCB-0527 Jy9-MC(B)"
		138 a 96 c 108 g 96 t
		BASE COUNT
		ORIGIN
		alignment_scores:
		Quality: 167.00 Length: 42
		Ratio: 4.175 Gaps: 1
		Percent Similarity: 95.238 Percent Identity: 80.952
		alignment_block:
		US-08-973-363-9 x BB830730 ..
		Align seg 1/1 to: BB830730 from: 1 to: 438
		1 lleleuProaspaspProsplylsProglinalalysglnleuglnth 17
		300 ATTCTCCAGATGATCCGTATAAAAACACAAACAAAGCTTACGAC 349
		17 rArgAlaAspTyrLeuIleTysLeuLeuAsnTysAspLeuAlaArgTys 33
		350 CCGTCGACGACTACCTCCCAACTAGCTAGACAGAGCTTGCAGAAAAGA 399
		34 luValGlnArgLeuThrGlyAla 41
		400 GAGGCTCAGAGCTTGTGTGTCGC 423
		seq_name: gb_estl:AL601246
		seq_documentation_block:
		LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001
		DEFINITION DKR2P31J1040.F1.313 (synonym: hlcc2) Homo sapiens cDNA clone
		VERSION AL601246
		KEYWORDS DKR2P31J1040.5, mRNA sequence.
		ACCESSION AL601246
		VERSION AL601246.1 GI:15164752
		KEYWORDS EST.
		SOURCE human.
		ORGANISM Homo sapiens
		REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
		1 (bases 1 to 430)
		AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, H. W., Well, B. and Wiemann
		S., (S. Bloeker, H., Boecher, M., Brandt, P., Mewes, H. W., Well, B. and Wiemann
		Wiemann, S.)
		COMMENT Unpublished (1999)
		CONTACT: Bloeker, H
		MIPS
		Am Klopferstr. 18a D-82152 Martinsried, Germany
		This is the 5' sequence of the clone insert
		Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
		Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
		Braunschweig/Germany) within the cDNA sequencing consortium of the
		German Genome Project.
		No sl sequence available.
		This clone (DKR2P31J1040) is available at the RZPD in Berlin.
		Please contact the RZPD: Ressourcenzentrum Heubnerweg 6, 14059
		Berlin-Charlottenburg, GFRMAN; Email: clone@rzpd.de.
		Location/Qualifiers
		1..430
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="DKR2P31J1040"
		/clone_lib="513 (synonym: hlcc2)"
FEATURES	source	Location/Qualifiers
		1..547
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:2443725"
		/clone_lib="NCI-CGAP_U02"
		/tissue_type="moderately-differentiated endometrial
		adenocarcinoma, 3 pooled tumors"
		/lab_host="DH10B"
		/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
		Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
		163 a 81 c 86 g 100 t
		BASE COUNT
		ORIGIN
		alignment_scores:
		Quality: 166.00 Length: 41
		Ratio: 4.256 Gaps: 1
		Percent Similarity: 95.122 Percent Identity: 82.927
		alignment_block:
		US-08-973-363-9 x AL601246 ..
		Align seg 1/1 to: AL601246 from: 1 to: 430
		1 lleleuProaspaspProsplylsProglinalalysglnleuglnth 17
		218 ATTCTCCAGATGATCCGATAAAAACACACAAACAAAGCTTACGAC 267
		17 rArgAlaAspTyrLeuIleTysLeuLeuAsnTysAspLeuAlaArgTys 34
		268 CCGTCGACGACTACCTCAATTAAGTGTAGATCTTGCAGAAAAGA 317
		34 luValGlnArgLeuThrGlyAla 41
		318 AAGCT.....CTTCTGTGTCGC 334
		seq_name: gb_estl:AI890775
		seq_documentation_block:
		LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000
		DEFINITION wmg5f11.x1 NCI-CGAP_U02 Homo sapiens cDNA clone IMAGE:2443725.37
		similar to: SM-CHD1_HUMAN.O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
		PROTEIN 1; mRNA sequence.
		ACCESSION AI890775.1 GI:5595939
		VERSION AI890775
		KEYWORDS EST.
		SOURCE human.
		ORGANISM Homo sapiens
		REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
		1 (bases 1 to 547)
		AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
		Tumor Gene Index
		COMMENT Unpublished (1997)
		CONTACT: Robert Strausberg, Ph.D.
		Email: cgsaps-r@mail.nih.gov
		Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
		Emmert-Buck, M.D., Ph.D.
		cDNA Library Preparation: Life Technologies, Inc.
		DNA Sequencing by: Greg Lennon, Ph.D.
		Clone distribution: NCI-CGAP clone distribution information can be
		found through the I.M.A.G.E. Consortium/LLNL at:
		www.bio.lnl.gov/bhbp/image/image.html
		Insert length: 1974 Std Error: 0.00
		Seq primer: -40up from Clpco
		High quality sequence stop: 418.
		Location/Qualifiers
		1..547
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:2443725"
		/clone_lib="NCI-CGAP_U02"
		/tissue_type="moderately-differentiated endometrial
		adenocarcinoma, 3 pooled tumors"
		/lab_host="DH10B"
		/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;

Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"  
BASE COUNT 114 a 118 c 85 g 230 t  
ORIGIN

## alignment\_scores:

Quality: 166.00 Length: 41  
Ratio: 4.256 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 82.927

## alignment\_block:

US-08-973-363-9 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

```
1 11leuPProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
378 ATTCCTCCAGATGATCCCGATTAACCAAGCAAAACAGTTGCAGAC 329
17 rArgAlaAspTYrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34
|||||
328 CCCTGCAGACTACCTCATCAATTAAGTAGAGATCTTGCAAAAAAG 279
34 luValGlnArgLeuThrGlyAla 41
|||||
278 AAGCT.....CTTCTGCTGCG 262
```

seq\_name: gb\_est1:AU125712

## seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,T., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isogai,T.)  
Unpublished (2000)

## JOURNAL COMMENT

Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

## FEATURES

## source

1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

## alignment\_scores:

Quality: 166.00 Length: 41  
Ratio: 4.256 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 82.927

## alignment\_block:

US-08-973-363-9 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

```
1 11leuPProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
450 ATTCCTCCAGATGATCCCGATTAACCAAGCAAAACAGTTGCAGAC 499
17 rArgAlaAspTYrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34
|||||
500 CCCTGCAGACTACCTCATCAATTAAGTAGAGATCTTGCAAAAAAG 549
34 luValGlnArgLeuThrGlyAla 41
|||||
550 AAGCT.....CTTCTGCTGCG 566
```

seq\_name: gb\_est2:BE895133

## seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1028)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-ri@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM9753 row: h column: 16  
High quality sequence stop: 488.

## FEATURES

## source

1. 1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPOrt6; Site:1; NotI;  
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

## alignment\_scores:

Quality: 166.00 Length: 41  
Ratio: 4.256 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 82.927

## alignment\_block:





Align seg 1/1 to: BF239967 from: 1 to: 821

```

6 ProAspLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22
   |||
3 CCGGATAAAAAACACAGCAAGCAACAGTTCACAGCCGTCACAGCTACT 52
   |||
22 uileuysleuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeu 39
   |||
53 CATCAATATCTAGTAGAGATCTGTGCAAAAAAGAAAGCT.....CTTT 96
   |||
39 hrcglyala 41
   |||
97 CTGGTGGC 104

```

seq\_name: gb\_estl:AM996787

seq\_documentation\_block:

LOCUS AM996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM996787  
VERSION AM996787.1 GI:8257021

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 337)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-230200-102-d03&ts=2000-02-23&cl=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

FEATURES

SOURCE

1..337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast; normal; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores: Quality: 129.50 Length: 35  
Ratio: 3.924 Gaps: 2  
Percent Similarity: 94.286 Percent Identity: 82.857

alignment\_block:

US-08-973-363-9 x AM996787/rev ...

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

```

1 lleuProAspLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 17
   |||
105 ATTCTTCAGATGATGCCGATTAACCAAGCAAGCAAGTGTGCAGAC 56
   |||
17 rargAlaAspTyrLeuileuys.leuLeuAsnLysAspLeuAlaArgLys 33
   |||
55 CCGTGAGACTATCTCATCAATCTT...AGTAGAGATCTTGCAAAAAA 9
   |||
34 Glu 34
   |||
8 GAA 6

```

seq\_name: gb\_estl:AM997058

seq\_documentation\_block:

LOCUS AM997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM997058  
VERSION AM997058.1 GI:8257292

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 686)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-150400-152-c03&ts=2000-04-15&cl=1)

Seq primer: puc 18 forward

High quality sequence start: 19

High quality sequence stop: 678.

FEATURES

SOURCE

1..686

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast; normal; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t

ORIGIN

## alignment\_scores:

Quality: 126.00 Length: 33  
Ratio: 3.818 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.879

## alignment\_block:

US-08-973-363-9 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

```
1 11leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
127 ATTCTTCGAGATGATCCCGATTA.AAACCAAGCAAAACATTCGAC 79
|||||
17 TATGAlaAspTyrLeuLeuLeuLeuAsnLysAspLeuAlaArgLys 33
|||||
78 CCCTGACAGCTACCTCTCATCTTAGTAGATCTTGCAAAAGA 30
```

seq\_name: gb\_gss:CNS04JPY

## seq\_documentation\_block:

LOCUS CNS04JPY 1047 bp DNA linear GSS 21-MAY-2000  
DEFINITION Tetradodon nigroviridis genome survey sequence T7 end of clone  
114P18 of library G from Tetradodon nigroviridis, genomic survey  
sequence.  
ACCESSION AL293839.1 GI:8032419  
VERSION AL293839  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetradodon nigroviridis.  
ORGANISM Tetradodon nigroviridis  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetradodon.  
1 (bases 1 to 1047)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bonneau,L., Billault,A., Queller,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradodon nigroviridis  
Unpublished  
2 (bases 1 to 1047)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,F., Brottier,F., Queller,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetradodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 1047)  
Genoscope.

JOURNAL Direct Submission  
TITLE Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases  
JOURNAL This sequence is a single read and was generated as part of a large  
COMMENT scale clone-end sequencing project of the Tetradodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradodon>.

## FEATURES

source  
1..1047  
/organism="Tetradodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="114P18"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COBGL1ADH09LP1-end : T7"  
BASE COUNT 226 a 270 c 244 g 297 t 10 others  
ORIGIN

alignment\_scores:  
Quality: 124.50 Length: 40  
Ratio: 3.662 Gaps: 1  
Percent Similarity: 85.000 Percent Identity: 65.000

## alignment\_block:

US-08-973-363-9 x CNS04JPY/rev ..

Align seg 1/1 to reverse of: CNS04JPY from: 1 to: 1047

```
1 11leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
465 ATTCTACAGATGATCCCGACAGAGCCCGATCCAGCAGCTACAGCG 416
|||||
17 TATGAlaAspTyrLeuLeuLeuLeuAsnLysAspLeuAlaArgLys 34
|||||
415 CAGAGCTGAGATATCTCTCAAGCTGCTGAAAAAGAACAGACAGACG 366
|||||
34 LuValGlnArgLeuThrGly 40
|||||
365 AG...CAGTCCCAKACAGCA 349
```





KEYTNPQIKQWRKNMIEVSKTFEDARKLHKYHAIKKROESQOHNDONISSNVN  
THVIRNDPDRLEKETTHNDSDSDSYSDHLSQYHDHNRDRGDAYKKSDSKRRPY  
SAFNGKDHMDHRYKQDSRYSDSKHRKLDHRSRDRSRLNLEGNLKDSGSHDRSH  
SDRHSDDHRSSTSEYSHKSSRDYRHSMDQMHRASGSGPRSPLDQRSYPGSSPLG  
HRSFHESSDKSTPEHTWSRKT"  
misc\_feature  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

alignment\_scores:  
Quality: 445.00 Length: 88  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 AspGluIleValSerValIysHisLeuHisLysLysIleLysThrGlu 17  
|||||  
4341 GATGAGATTGTTTCAGTGAAACATCTACATATAAAAAATAAAAACAGAAAA 4390  
17 sGluAsnGluGluLysProGluProAspIleGlyIleLysLysGluVala 34  
4391 AGAAATATGAGAAAAAGCCGACCCAGATTTGCTATTAAGAAAGAGACG 4440  
34 LuGluLysArgGluThrLysGluLysGluAsnLysArgGluLeuLysArg 50  
4441 AAGAAAAAGAGAGACAAAGAGAAAGAAAAATAAAAAGCAATGAAAAAG 4490  
51 GluLysLysGluLysGluAspLysLysLysGluLeuLysGluLysAspAsn 67  
|||||  
4491 GAGAAAAAGAAAAAGAGATAGAAAGATTAAGAAAAAGATTAATAA 4540  
67 sGluLysArgGluAsnLysValLysGluSerThrGlnLysGluLysGlu 84  
4541 AGAAAAAGAGAAAAACAAGTAAAGAAATCCACACAGAAAAAGAAAAAG 4590  
84 aLysGluGluLys 88  
|||||  
4591 TCAAGAGAGAGAGAG 4604

seq\_name: gb\_pat:A58693

seq\_documentation\_block: 265 bp DNA linear PAT 06-MAR-1998

LOCUS A58693  
DEFINITION Sequence 12 from Patent WO9639505.  
ACCESSION A58693  
VERSION A58693.1 GI:3714251  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.

REFERENCE 1 (bases 1 to 265)

AUTHORS Griffiths,R. and Tiwari,B.

TITLE AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN

JOURNAL BIRDS  
Patent: WO 9639505-A 12 12-DEC-1996;

COMMENT ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.

FEATURES location/Qualifiers  
1..265

BASE COUNT 158 a 16 c 61 g 30 t

ORIGIN

alignment\_scores:

Quality: 432.00 Length: 89  
Ratio: 4.909 Gaps: 1  
Percent Similarity: 98.876 Percent Identity: 98.876

alignment\_block:

US-08-973-363-11 x A58693 ..

Align seg 1/1 to: A58693 from: 1 to: 265

1 AspGluIleValSerValIysHisLeuHisLysLysIleLysThrGlu 17  
|||||  
1 GATGAGATTGTTTCAGTGAAACATCTACATATAAAAAATAAAAACAGAAAA 50  
17 sGluAsnGluGluLysProGluProAspIleGlyIleLysLysGluVala 33  
51 AAGAAATATGAGAAAAAGCCGACCCAGATTTGCTATTAAGAAAGAGACGT 100  
34 GluGluLysArgGluThrLysGluLysGluAsnLysArgGluLeuLysArg 50  
101 GAAGAAAAAGAGAGACACAAAGAGAAAGAAAAATAAAAAGCAATGAAAAAG 150  
50 gGluLysLysGluLysGluAspLysLysLysGluLeuLysGluLysAspAsn 67  
151 GGAGAAAAAGAAAAAGAGATAGAAAGATTAAGAAAAAGAAAGATTAATA 200  
67 sGluLysArgGluAsnLysValLysGluSerThrGlnLysGluLysGlu 83  
201 AAGAAAAAGAGAAAAACAAGTAAAGAAATCCACACAGAAAAAGAAAAAGAA 250  
84 VALysGluGluLys 88  
|||||  
251 GTGAAGAGAGAGAGAG 265

seq\_name: gb\_pr:AC092372

seq\_documentation\_block:

LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
ACCESSION AC092372  
VERSION AC092372.3 GI:17402768  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 101220)

AUTHORS Doe Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 101220)

AUTHORS Doe Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.  
Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
www.sngc.stanford.edu

Finishing Completed at Stanford Human Genome Center

Quality: Phrap Quality >=40 100% of sequence;  
Estimated Total Number of Errors is 0.

NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146.7kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.

FEATURES location/Qualifiers  
1..101220  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT	34122 a	18862 c	17827 g	30409 t	/chromosome="5" /clone="RP11-58M12"
ORIGIN					
Alignment_scores:					
Quality:	356.00				Length: 88
Ratio:	4.395				Gaps: 0
Percent Similarity:	92.045				Percent Identity: 78.409
alignment_block:					
US-08-973-363-11 x AC092372/rev	..				
Align seq 1/1 to reverse of: AC092372	from: 1 to: 101220				
1	AspGluIleValSerValIleuHisLeuHisLysLysIleLysThrGluIly 17				
25410	GAAGAAATCTAGTCTGTGAAACATCTCAATAAAAAAATTTAAACAAAG 25361				
17	sgLysAeGluGluLysProGluProAspIleGlyIleLysLysGluLysAag 34				
25360	AGCAATGAGAGAAAAACCTGACCCAGATGTTATATTAAGAGAGAACG 25311				
34	IuGluLysArgGluThrLysGluLysLysLysArgGluLeuLysArg 50				
25310	AAAGAAAGAGCGAAGCAAGAAAAAGAGATTAATAAAAGCTTAAGG 25261				
51	GluLysLysGluLysArgLysLysLysLysLysLysGluLysAspLys 67				
25260	GAGATTAAGAAAAAGAGATTAAGATTAAGAAAAAGATTTTA 25211				
67	sgLysArgLysLysLysValLysGluSerThrLysGluLysGluLys 84				
25210	AGAAAAAGAGAAAAAGAAAAAGATTAAGATTAAGAAAAAGCA 25161				
84	allysGluGluLys 88				
25160	TAAAGAGAAAAAG 25147				
seq_name: gb_pr:AC012624					
seq_documentation_block:					
LOCUS AC012624	134365 bp	DNA	linear	PRI 21-JUL-2001	
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.					
ACCESSION AC012624					
VERSION AC012624.6	GI:14993679				
KEYWORDS HTG.					
SOURCE					
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 134365)				
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.				
JOURNAL	Direct Submission				
REFERENCE	2 (bases 1 to 134365)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
REFERENCE	3 (bases 1 to 134365)				
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
REFERENCE	4 (bases 1 to 134365)				
AUTHORS	Direct Submission				
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.				
JOURNAL	Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
COMMENT	On Jul 21, 2001 this sequence version replaced gi:14277267.				
FEATURES	Location/Qualifiers				

```

source      1..134365
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2082117"
BASP COUNT   40414 a 24497 c 25503 g 43951 t
ORIGIN
alignment_scores:
    Quality: 356.00          Length: 88
    Ratio: 4.395           Gaps: 0
    Percent Similarity: 92.045     Percent Identity: 78.409
alignment_block:
US-08-973-363-11 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365
1 AspgtluilevalserValysHisleuHisLysLysIlellysThrGly 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119773 GAAGGATCAGTCTGTGTAACATCCAAATCAAAAAAATTAAAACGAAG 119822
17 sgtuasngluGlulysPrGluPProAspIleGlyIlellysSGUUAAG 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119823 AGACAGTGTAAGAAAAACTCGAGCCAGTGTATATATAAAGAACAACAG 119872
34 luGlulysArgGluThrLysGluLysLysLysArgGluLeuLysArg 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119873 AAGAAAAAGAGGCAAGCAAGAAAGAAAGAGATATAAAAAGAACTTAAAG 119922
51 GlulysLysGluLysGluAspLysLysGluLeuLysGluLysAspNly 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119923 GAGATATTAAGAAAAAGCATATACAAAGATATTAAGCAAAAAAGATTTTAA 119972
67 sglulysrsgLusnrlsyalysGluSerThrGlnLysGluLysGluV 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119973 AGAAAGAAAGAAACAAAGATTAAGAAAGCTATACAGAAAGAAAGCA 120022
84 allysgLugLulys 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120023 TAAAGAGAAAAAAG 120036
seq_name: gb_htg.AC021449
seq_documentation_block:
LOCUS       AC021449               143079 bp        DNA             linear      HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION   AC021449
VERSION     AC021449.3 GI:10047806
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 143079)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-58M12
JOURNAL     unpublished
REFERENCE   2 (bases 1 to 143079)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckert,Y.R., Bedalov,A., Boguski,M.S., Bouknight,R.B., Brown,A., Burkett,D.G., Castle,A., Choquel,I.Y., Colangelo,K., Collins,S., Collumore,A., Cooke,P., DeRubeis,K., Dewar,K., Domingo,M., Doyle,M., Fencsik,R., Ferrerella,P., FitzHugh,W., Forrest,C., Gaye,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howard,J.C., Johnson,S., Jones,C., Kamp,L., Karalas,A., Klein,J., Landers,T., Lehocsky,J., Levine,R., Lien,C., Liu,G., Locke,K., McDonald,P., Margulis,N., McKean,P., McGuck,A., McKernan,K., Mcpheeters,R., Meldrum,J., Menous,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,

```

TITLE  
JOURNAL  
COMMENT

Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange, Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J., Threlle, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wymen, D., Ye, W.J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L5154  
Center clone name: 58\_M\_12

----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 14400; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp
* 40512 43279: contig of 2768 bp in length
* 43280 43379: gap of 100 bp
* 43380 46905: contig of 3526 bp in length
* 46906 47005: gap of 100 bp
* 47006 51830: contig of 4825 bp in length
* 51831 51930: gap of 100 bp
* 51931 62619: contig of 10689 bp in length
* 62620 62719: gap of 100 bp
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp
* 75509 92516: contig of 17008 bp in length
* 92517 92616: gap of 100 bp
* 92617 106409: contig of 13793 bp in length
* 106410 106509: gap of 100 bp
* 106510 143079: contig of 36570 bp in length.
```

## FEATURES

## Source

## misc\_feature

1. .143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_1lb="RPCT-11 Human Male BAC"

## misc\_feature

/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
38921. .40411  
/note="assembly\_fragment"  
40512. .43279  
/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

```
misc_feature 43380. .46905
/note="assembly_fragment"
misc_feature 47006. .51830
/note="assembly_fragment"
misc_feature 51931. .62619
/note="assembly_fragment"
misc_feature 62720. .75408
/note="assembly_fragment"
misc_feature 75509. .92516
/note="assembly_fragment"
misc_feature 92617. .106409
/note="assembly_fragment"
misc_feature 106510. .143079
/note="assembly_fragment"
misc_feature clone_end:R7
vector_side:right"
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others
ORIGIN
```

alignment\_scores:  
Quality: 356.00 Length: 88  
Ratio: 4.395 Gaps: 0  
Percent similarity: 92.045 Percent identity: 78.409

alignment\_block:  
US-08-973-363-11 x AC021449 ..

Align seg 1/1 to: AC021449 from: 1 to: 143079

```
1 AspGluIvalSerValIysHisLeuHisIysIleYsThrGluIly 17
||||||| ..|||||||
117669 GATGAGATCAGTCTGTGAACATCCAAATATAAATAAAGCAGAAAG 117718
```

```
17 SGLuSnGLuGluIlyProGluProAspTleGlyIleIysIysGluAlaG 34
||||||| ..|||||||
117719 AGACAGTGAAGAAACCTGACCGATGTTATATTAAGAAAGAACCG 117768
```

```
34 LuGluIysArgGluThrIysGluIysGluAsnIysArgGluIysArg 50
||||||| ..|||||||
117769 AAGAAAGAGGAGCAAGCAAGAAAGAGAAATTAAGAAAGAACTTAAAGG 117818
```

```
51 GluIysIysGluIysGluAspIysIysGluIleuIysGluIysAspAsnIly 67
||||||| ..|||||||
117819 GAGATAAAGAAAGAAAGAGATTAAGAAAGATTAAGAAAGAAAGATTAA 117868
```

```
67 SGLuIysArgGluAsnIysValIysGluSerThrGluIysGluIysGluIy 84
||||||| ..|||||||
117869 AGAAAAAGAGAAACAAAGTAAAGAAAGCTATACAGAAAGAAAGACAA 117918
```

```
84 alIysGluGluIys 88
:::|||||||
117919 TAAAGCAAGAAAG 117932
```

seq\_name: gb\_hlg:AC008531

seq\_documentation\_block:

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.

ACCESSION AC008531 GI:12830078  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 145659)  
DOE Joint Genome Institute.

## AUTHORS

Sequencing of Human Chromosome 5  
Unpublished

## REFERENCE

2 (bases 1 to 145659)  
DOE Joint Genome Institute.





```

* 9097 9196: gap of unknown length
* 9197 10260: contig of 1064 bp in length
* 10261 10360: gap of unknown length
* 10361 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length
* 12561 14611: contig of 2051 bp in length
* 14612 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40979: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 67981: gap of unknown length
* 67982 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.

```

FEATURES

source

```

1. 193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-36012"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN

```

alignment\_scores:

Quality:	Length:
356.00	88
Ratio: 4.395	Gaps: 0
Percent Similarity: 92.045	Percent Identity: 78.409

alignment\_block:  
us-08-973-363-11 x AC091946 ..

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 AspGluIleValSerValIysHisLeuHisIleValIleValThrGluIy 17

89180 GATGAGATCAGTTCTGTGTAACATCCAAATATAAATAATTAACGAAAG 89229

17 sGLuAsnGluGluIysProGluProAsp11eGlyIleIysIysGluAlaG 34

89230 ACACAGTGAGAGAAAACCGACCCAGATGTTATATTAAGAAAGAACAG 89279

34 IuGluIysArgGluThrIysGluIysGluAsnIysArgGluIleIysArg 50

89280 AAGAAAGAGCGACAGCAAGAAAGAAAGAAAGAAAGAAAGAACTTAAAG 89329

51 GluIysIysGluIysGluAsnIysGluIleIysGluIysAspAsnIy 67

89330 GAGATTAAGAGAAAAGAGATTAAGATTAAGATTAAGAGAAAAGATTTTAA 89379

67 sGluIysArgGluAsnIysValIysGluSerThrGlnIysGluIyuv 84

89380 ACAAAGAGAGAAACAAAGTAAAGAGCTTACAGAAAGAAAGAAAGCA 89429

84 aIlysgIuGluIys 88

89430 TAAAGCAAGAAAG 89443

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778

VERSION AC026778.4 GI:14277282

KEYWORDS HTG.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 195433)  
DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 195433)  
DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission  
Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 1, 2001 this sequence version replaced gi:13677045.  
Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center

www.sbgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence:

Estimated Total Number of Errors is 0.2.

STS Content:  
WI-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841.

FEATURES

source

```

1. 195433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-428111"

```

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN





[illegible]

```

Alignment_scores:
  Quality: 189.50      Length: 79
  Ratio: 2.961        Gaps: 3
  Percent Similarity: 81.013      Percent Identity: 55.696

Alignment block:
  US-08-973-363-11 x AL589701/rev ..

Align seg 1/1 to reverse of: AL589701 from: 1 to: 219200

11 LyslysllelysthrtglutylsglaasglutglutylsProoLuproAsp11 27
|||||
|||||
|||||
|||||
|||||
61334 AAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG.. 61277
27 eglyllyelyslgclualaglutglutylsArgclutHrlysgllysglUA 44
.....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 61234
44 snlysatrgLlueulyArgclutylsLysglulysglUAAspLylysglu 60
|||||
|||||
|||||
|||||
|||||
61233 AGAAGAGAGAG...AAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 61187
61 LueulysgLlysaAspAnlysgLlysaArgcluaAnlyValylsgluSe 77
|||||
|||||
|||||
|||||
|||||
61186 AAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 61137
77 rThrgllyys...glulysglUVallysgclutglutyls 88
:|||||
:|||||
:|||||
:|||||
:|||||
61136 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61100
seq_name: gb_htg:AL645746

seq_documentation block:
LOCUS AL645746 282611 bp DNA linear HTG 12-JAN-2002
DEFINITION Mus musculus chromosome 13 clone RP23-153B6, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL645746
VERSION AL645746.2 GI:17148444
KEYWORDS HTG; HTGS_PHASEL.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.
1 (sites)
Sims, S.
Direct Submission
Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humuqres@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
on Nov 28, 2001 this sequence version replaced gi:17017892.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: bm153B6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: piasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 279809 bases at least Q40
Consensus quality: 280516 bases at least Q30
Consensus quality: 281062 bases at least Q20
Insert size: 281711; sum-of-contigs
Insert size: 190968; 4.7% error; agarose-fp
Quality coverage: 12.3% in Q20 bases; sum-of-contigs quality
coverage: 18.52% in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence

```





/protein\_id="Aac58053.1"  
/db\_xref="GI:2337969"  
/translation="MSQNSNSNENSPKRRKYVKMCDLFEOKERRRSINBRASKNFLK  
RRRIPEOQOEKGLINLKYSNRRCOVERKRDRLIRLWNLKCTONNYNGPPE  
PRYKVENSLMOCATAFNLIDQOTTNINLIPEVSGNNNTNGFAATATLHTNCKEK  
TLANNINNFKAQNLCEVLPFSFTSALDULSIDMWNLYNL"  
3492..5453  
/note="A3: AHV-sema, similar to Vaccinia A39"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="semaphorin homolog"  
/protein\_id="Aac58054.1"  
/db\_xref="GI:2337970"  
/translation="MAYLNATVSKPVLISLSKYLKFEHCGEGCGLITTEFVH  
PAMGTCVSIIRILMILATAKSRFDKRLIVNLTDGGOHREFGPOQPHVLFH  
SLNSDVGNGNNTIYLFDFHSSNASTALINTSTPHRLSTGCNPTTLIHQND  
GLACGNSOKPSCWILNNITTOPLGKCLALPSPSGNLYLFDQNDITSTNLKYS  
LGSCHKRRIRIAGVVELTSDTAHRRPOVATVHKNESYDCKIYFFQENSHSDEKQ  
EPHPTVPVGVQSSDQGESLSVYKMTTFELKARLACVDYDGRKIVNELQDIFIMWQ  
ENSMETLIYGLFLSPWNFSACVFTVKDIDHFKTSKLKNHHKLPTRPGQCMKH  
QHVPTFQVADRYREYADPVYOKNNAMPILIOSKYITKLAVRYEGGVNATIFY  
LMTIKGTHIVYREDSNSTALILEINFORPAPIONITLNTNKLKLVNEMEVS  
EVPIDLCYVGNOCFCFMSRDPCTIYNNTCSEFKQVSEYETGPNARTISEMGDIY  
APTVKHQVSIPLISNYSLSPVSNMADFWTKDFTFKCHVKTKHKNCCIILIAN  
TTATNGHVCNMKEDSVTKLEVNVTLM"  
5590..5597  
5732..6097  
/note="ORF containing putative signal peptide; detected by  
5'RACE-PCR"

polyA\_signal  
CDS

/codon\_start=1  
/evidence=experimental  
/product="A4"  
/protein\_id="Aac58055.1"  
/db\_xref="GI:2337971"  
/translation="MVAOLYHHLTILIASIYIIFVNAAPTLXEDDEPNSTEYDF  
LSDSRQFPRNNSCNCRTREARTRTLLEALKRALQIITNETIDELPTPTMTT  
PLETTPDTSPPVPSAIP"  
6123..6128  
/evidence=not\_experimental  
6333..10442  
/note="ORF03: similar to H. salmirci and EHV2 ORF3, similar  
to ORF75"

polyA\_signal  
CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="tegument protein"  
/protein\_id="Aac58056.1"  
/db\_xref="GI:2337972"  
/translation="MITREQLLAATRAEPALSTAAIAIRSTPGDQNYITQDAVL  
VPTFSRPRRLREHOLKIETLFTKALRTIDEQOYLPTLSALSSSTFYVGPDIINL  
PTSQSNELTALIQTRADAREKVENVERIOHCRLLFFSGSGEQLNHTHYALLREILCD  
LTLYOFRNELFNSLDFYATIYPRFLKDKVRHYGIDINLTGAASNYNASPT  
AVDSEIFRIPYHVVRFIDRWEIRVYRPPVGVGHSRVRQHGGAAYNSTLNSK  
GATGLOKGMITSFISOPTGOLGILTEPFGMGQDELSTATIOMLTIKVNTYBAPK  
TLNAYGVPLVQGVFVLSPTLREKRAAFGTSLSITGPIYILPPTAETIDAMERRK  
OULIVEGYRPIYDGKGLHNPVNTNNSSGRHAIHLHOLQMLITPTEGVVHICDM  
QNEKVTMRKIAACGDLGSMARSRLPAHTLRLKKNRNRLNYYNRITIKNSLWQ  
SAALVVIADNEOVDSKMDIALSGGCEPHLIGTLEDLLQVLRHPVGGKANHVVH  
OYKMHMPKQSEGETEDPLAPDSNIQLEKNLDTEDLLQVLRHPVGGKANHVVH  
DRCGNGHIAQGVGPEDIPICDESVTVHNLVDGDI REGMESVRYMAADRYARALI  
ERYGTPGSDITPATLANNIGLHYHHPSEQVYVESKRYCNGCIGIKTTFTQBDPL  
GITLAIYESCTNCTIGVENTEELIGLSTVPEGIHYRREHNSIMAAADFCSSNMF  
GQVNSAENENCLRSVYATANAPCVVPGPSLPYFKKPGSALIRVNLHNEHLSGGI  
CMAASIGASEETPTPSOLANLQLEMLYKARNALSGHVSADGGLICAVCEMFG  
GLSARLIHDEDEPVPLESETPGFLEVNALIDVAILIRANLYNECQIOGHEVYS  
DFTFVHONTOILSVPSRLKNHMTLEFSKVDLIYKEDOVLPBETSQYVNEHLYTD  
PYSLISOSTRPNVLYHLPGCGT PDLALALNSGSPDTVYIPGCKIYHNRKEDAA  
PGSPYADDFIAGIYLVIGSSNIDSDVGSTIRQMLNVRQYINDVRRNLKSGSTJAL  
GOLACRLIFATKAIGFDAGSQPTFLPNASRYESRWLNFKIPEDTKAFAFDRLAC  
VLPCWVGTHIGFSHNNTIPEGDLERQOAAATGNGPLVOSGAPAREPLAPTEAEHRY  
AGCSEDEGRLALLPDCLAFTNMQHNQOTGPGGGLPYSPAKIMRYRLYNMSKFFHQ  
HYRSLRTNLRHNFNETQOQDLDFPHADRRAVHDNDQYITPMDAPSO"  
11134..11139  
/evidence=not\_experimental

polyA\_signal

CDS

11260..14643  
/note="ORF06"  
/codon\_start=1  
/evidence=not\_experimental  
/product="major ss DNA binding protein"  
/protein\_id="Aac58057.1"  
/db\_xref="GI:2337973"  
/translation="MALKHQLNHAEVNDLNSKAPIGCGFIYIYPTHEFKETSLIG  
DKYAEAGFSLPLIHGVTEAFVNPVNAKAYKKIDMTYTSVKISTSTYNAIIPHNK  
FESTFSGRGISLCKEACDFGVYPTPILEGESGTSVSDICPPYQOEKDALMAYVITBG  
FKERHLGKLIYILKQMSHVMINKTEYRIRLPLDEDLTKSSLSRLRLYPAVSEYLY  
TLVYLSAQSLRVHNAASLVEAIQEOPHDXYKAKLVSEFVEYLAIVAGADDTLMDI  
AVAAELGLSYSLSEFFAPQEKTRVQDYSDVASETQSDREALSALMAQALAIH  
AOLPSESIYYVNRVAAOAPIPSKNEPVYNSYVIOHGLANLCEETLEFDGSPAFG  
APASLIDGSEFTLOHLAAAFSPNLLABWCYALQOCROKSTLNAMNYCTEVGSA  
NSPVCSLCSQCCVCYCNITFYALKQRFPPVLDGSRDPYVITGTTNVEDELFGNF  
ASFRDKEDONQTEETPRYITWOLNQTLEKLAAGVLDSPVADBEAGSGSGSNLEKF  
VRPEFSDLSLVDAEAKETMTMKNNVNFKESIKGVSHVIOYKNYVQWAPCSLMLNL  
YYSILITIIODIALPISTYSESNPAGQYKPNMKLIHQTMTNFKSEFIDKGVITG  
TEKMYVABQSPDFEDVATNNMYSPVQVRLARAQVATLKNIKVKNRILRSGTSM  
SEHYQNAFLKTANRQNYILAGPYVFLANSFHRQLPMLKISCLYIAMSNCRCQIOP  
VPGYSALAKKPFYSYINNSKQBEVNMADVPDYVYTTAKQOLNALLIAPCQOTQIP  
AVTHTSIFPVQETCALEYPIVGLTSSVDSVEDYVNNVQNLKALTVNSSLRESANIDA  
RSRPVITLPPVYVNRKYTGIAQNAQLFQSANLGYPMGRGVQNDLIGDSLFPKKQNSYMR  
KYLIFMTPLVGNLKPSTYHQGAFFELVYKRTIOSLIDQADEVDYLNRYVCELVS  
GACCAEDLTDDIOFYLGSGYGFSENIIEKIDQIRELVGPYTHMAESVYLSKGTCEPDE  
VOEFAFEEOVKRLTSMDSHSGYVGGKKRKIATMPDDLTL"  
14725..16767  
/note="ORF07"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="transport protein"  
/protein\_id="Aac58058.1"  
/db\_xref="GI:2337974"  
/translation="MGQYLAALYQIYGLCDVSYLVEPCKPTSLCTKTIADACNKH  
IHEVSAHSLLQONSLEKCALSTLSHLEMLKRLFFYTHALDNPYSLKSSVGL  
CDLHAKKLNWQFVNEGCIENVLTLINIERLSLNCVFTYALDNPYSLKSSVGL  
QLRGISEVPERTDIYITSSCLCEVLFTSVYVNGOETLNEILNHNHNLHYERVPEPI  
KGLFESELQNLGLKVAHATDITIOSGVKEHVAQESLVAIKATITNNPKQVLELNP  
LLYVNSGONQSPDSGVKCSSELKIMSNEMLQKRYRKLNNLEPPHFDLHSGTREL  
LPGGIFSTPHDTALKQDCSNMPEKQPTLVAKRONLLEPRLSNIILYGEVPPKP  
KOTBSALKTCDQSDASKNOYLOEAELEKRAYLKLKSEGRKLOACSTHEBNLSQ  
SLKIVGSVYVYKQSATLLNHLFQRSWVTOQSLPVSNGSPQESNFKISLSLYVSL  
SRBYLSTRLHAFPALITGPLTTOBGLFPPSPNOVLACHAEHMPHOKMLINEMKP  
HMPQDMICSNFNEFYTIHETDLNGVYECWKILRELYLSVALYNTWEKNLCTIYRD  
TSCPTQACSSGIRKGLVYTESHAPLILVYSNKKWITPKDIALLYAHMQLANNAHR"  
15048..15053  
/evidence=not\_experimental

polyA\_signal  
CDS

/note="ORF08: similar to EBV BALF4, CMV UL55, HSV UL27"  
/codon\_start=1  
/evidence=not\_experimental  
/product="glycoprotein B"  
/protein\_id="Aac58059.1"  
/db\_xref="GI:2337975"  
/translation="MAHTGSTVCAFLIFAVLKNVFCQPTSSSEVEDVDEPANTVSDN  
IIRQORNTKCGIHSBPAPFRVCSANIGDIFRQTSHCNPTDKENESILLIF  
KENIVYVFEVRYKRYKIYVTSITLYNGIYADAVANOHVESVPIYERBMDITYOCN  
SLDVTYGNGLIYVTDNDGSMYTDIDLPQVGLSSVRYRHSQPIIAHEPMLGLGYYRR  
TYNCEVETEDANAVPEFRIFTINIGDTIEKSPFSANMETEFGSDPDTLIVAKQY  
RVNDYKERGQPOGHRIFEDKEEYTLISMAQOPRNTSYCMAMWKSFDNAIKTEHGS  
LHNVANDITASFYTPPTQREVGLHVCNLITIESELKSLANVNDTHSPNGAQVYL  
TNGGLLVMOPLAOOKLMDAKGLDVAKQONNTTTTTRRSORRSVSGGDVYLT  
AESTIITLITQIOPAYDRLQINNVLIELISAMKREOHRAHNSIMNELSKINPTVMSI  
YGRPVSAKRIQGVIVSYSHCVYVQDVSILRSKRVPRQDKTHCYSRPPTFFKFNDS  
HLKRGQGVNNELITLTTVAETICHENHTEFQCGNNMYEKNRHVKTMYGVATLID  
TFWNLITLIVENIDFOVIELYSREBKMSAFPIELIMFERYNYTQVGLRQGLDMLFG  
ATNRNDFVDAGSLMDIDGVYATLVNASSVATLSSISYGIILNLIKNFQGMILLFG  
LIAVAATVTLIRKAKKRFQONQVOMIYPIKIKITTSORELOVDPSIKHBLDITLAM  
HDVHASQPSKQDEDEGSTTSQAPMLNKAKNVLRRAQYKRLKRTDSESGVY"  
19314..19319  
/evidence=not\_experimental

polyA\_signal  
CDS









	1	GATGAGATTGTTCACGTGAACATCTCACTATAAAAAAATAAAAACAAGAAA	50
	17	ysgluabnsgluylusProgluproaspilleglyilelysLysgluala	33
	51	AAGAAATATGAAGAAAAAGCCTGACGCAGATTTGGTATPAAAGAAGAGACT	100
	34	GlUGlUlyArGltwThrLysGLuLSLuSnLySArGtULeulyAr	50
	101	GAGAAAAAAGAGAACAAAAAGAGAAATAATAAACGCAATTGAAAG	150
	50	gGUlySLySGlUlySGluAsPLySLyGLuLeuLySGlUlySAsPaNL	67
	151	GGAGAAAAAAGAAAAAGAGCATTAAGAAACAAATTAAAAAGAAAAAATATA	200
	67	yseGluLySArGluSnLySAlySLyGluSeThrGlnLySGlUlySGlu	83
	201	AAGAAAAAGAGAAAAACAAGTAAAGAAATCCACACAGAAAGAAAAAGAA	250
	84	VAllySGlUGlUlyS	88
	251	GTCAGAGAGAGAGAG	265
seq_name:	/SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA1997.DAT.AAT42753		
seq_documentation_block:	ID AAT42753 standard; cDNA; 137 BP.		
XX	AAT42753;		
XX	AC		
XX	DY		
XX	12-MAR-1997	(first entry)	
XX	DE		
KW	Chicken CHD-W clone CC14 3' motif.		
KW	Bird; sex determination: chromodomain-Helicase-DNA binding 1 Avian;		
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.		
XX	Gallus sp.		
XX	OS		
XX	Key	Location/Qualifiers	
FT	misc_difference 52	/*tag= a	
FT		/note= "base 52 disrupts the reading frame for	
FT		the translated amino acid sequence given	
XX		in Fig 7"	
PN	WO9639505-A1.		
XX	PD		
XX	12-DEC-1996.		
XX	PF		
XX	05-JUN-1996;	96MO-GB01341.	
XX	PR		
XX	06-JUN-1995;	95GB-0011439.	
XX	PA		
XX	(ISIS-) ISIS INNOVATION LTD.		
XX	PI		
XX	Griffiths R, Tiwari B;		
DR	WPI: 1997-043127/04.		
XX	P-FSDB; AAW08145.		
PT	Avian chromodomain-helicase-DNA binding genes determine sex in		
PT	birds - used for sex determ. and to control sex of progeny		
XX			
PS	Disclosure: Fig 7; 76pp; English.		
XX			
CC	A composite (incomplete) nucleotide sequence (AAT42753) and putative		
CC	translation (AAW08145) sequence are provided of a motif that is found		
CC	spliced to the 3' end of chicken CHD-W clone CC14 (see also AAT42754)		
CC	There are no splice donor or acceptor sites within the motif		
CC	suggesting it is a final rather than an intermediary product of		
CC	splicing. The motif is also found as an insert in some CHD-IA clones		
CC	(see also AAT42752).		

```

XX      SQ      Sequence 137 BP; 75 A; 10 C; 34 G; 18 T; 0 other;
XX
XX      alignment_scores:
XX          Quality: 180.00          Length: 46
XX          Ratio: 4.390          Gaps: 1
XX          Percent Similarity: 89.130          Percent Identity: 84.783
XX
XX      alignment_block:
XX      US-08-973-363-11 x AAT42753          ..
XX
XX      Align seg 1/1 to: AAT42753 from: 1 to: 137
XX
XX      1 AspGluIleValSerValIysHisIleuHisIysLysIleIysThrGlu.L 17
XX      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      1 GATGGATGTTGTTCCAGTGAACATCCACATATATAAATAAAGCAGAAAA 50
XX
XX      17 ysgLusngIugIuLysProGluProAspIleGlyIleIysLysGluAa 33
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      51 AAGAAATGTAAGAAAAAGATGAGCCAGATGTTGTTATTAAGAAAGAAAGCT 100
XX
XX      34 GluGluLysArgGluThrIysGluLysGluAsnLys 45
XX      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      101 GGAGAAAAAAGAGACACAAAGAGAAAGAAATATAG 136
XX
XX      seq_name: /SID1/gcgsdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS90688
XX
XX      seq_documentation_block:
XX      ID AAS90688 standard; CDNA: 621 BP.
XX
XX      AC AAS90688;
XX
XX      DX 13-FEB-2002 (first entry)
XX
XX      DE DNA encoding novel human diagnostic protein #26492.
XX
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX      food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX      Homo sapiens.
XX
XX      OS
XX      WO200175067-A2.
XX
XX      PD 11-OCT-2001.
XX
XX      PE 30-MAR-2001; 2001WO-US08631.
XX
XX      PR 31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX
XX      PA (HYSE-) HYSEQ INC.
XX
XX      PI Dmanac RT, Liu C, Tang YT;
XX
XX      WPI; 2001-639362/73.
XX      P-PSDB; ABG26501.
XX
XX      PT New isolated polynucleotide and encoded polypeptides, useful in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits and to assess
XX      biodiversity -
XX
XX      Claim 1; SEQ ID No 26492; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      (II). (II) is useful for generating antibodies against it, detecting or

```











CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 297 BP; 153 A; 12 C; 112 G; 20 T; 0 other:

## alignment\_scores:

Quality:	155.50	Length:	81
Ratio:	2.681	Gaps:	2
Percent Similarity:	71.605	Percent Identity:	44.444

## alignment\_block:

US-08-973-363-11 x AAS90725 ..

Align seg 1/1 to: AAS90725 from: 1 to: 297

```
8 H1sLeuH1sLysLys.....11eLysThrGluLys 17
|||||:|||||
59 CACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 108
|||||:|||||
17 sgluasnglulysProgluProaspIleGlyIleLysLysGluLac 34
|||||:|||||
109 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 146
|||||:|||||
34 luclulysarGluThrLysGluLysGluLysarGluLysarG 50
|||||:|||||
147 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 196
|||||:|||||
51 GlulysLysGluLysGluLysLysLysLysLysLysLysLys 67
|||||:|||||
197 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
|||||:|||||
67 sglulysarGluLysLysLysLysLysLysLysLysLys 81
|||||:|||||
247 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
|||||:|||||
```

seq\_name: /SIDSL/9csgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.ABA58819

## seq\_documentation\_block:

ID ABA58819 standard; DNA; 475 BP.

XX ABA58819;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #7124.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0633366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human foetal liver -

XX Claim 1; SEQ ID NO 7124; 639pp + sequence listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 475 BP; 38 A; 198 C; 45 G; 194 T; 0 other:

## alignment\_scores:

Quality:	155.50	Length:	83
Ratio:	2.430	Gaps:	2
Percent Similarity:	77.108	Percent Identity:	42.169

## alignment\_block:

US-08-973-363-11 x ABA58819/rev ..

Align seg 1/1 to reverse of: ABA58819 from: 1 to: 475

```
7 LysH1sLeuH1sLysLysLysLysLysLysLysLysLysLys 23
|||||:|||||
364 AAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 315
|||||:|||||
23 ogLupProaspIleGlyIleLysLysGluLacLysLysLysLys 40
|||||:|||||
314 AGAGAGAG.....AAAGAGAGAGAGAGAGAGAGAGAGAG 277
|||||:|||||
40 ysgLulysGluLysLysLysLysLysLysLysLysLysLys 56
|||||:|||||
276 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227
|||||:|||||
57 AspLysLysGluLysLysLysLysLysLysLysLysLysLys 72
|||||:|||||
226 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
|||||:|||||
72 nlYsVallysgLysLysLysLysLysLysLysLysLysLys 88
|||||:|||||
176 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128
|||||:|||||
```

seq\_name: /SIDSL/9csgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.ABA27737

## seq\_documentation\_block:

ID ABA27737 standard; DNA; 475 BP.

XX ABA27737;

XX 23-JAN-2002 (first entry)

XX Probe #6203 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX

[illegible]

```

XX Human: brain expressed exon: gene expression analysis: probe:
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX W0200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 6964; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Seq
XX
XX Sequence 475 BP; 38 A; 198 C; 45 G; 194 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 155.50      Length: 83
XX      Ratio: 2.430      Gaps: 2
XX      Percent Similarity: 77.108      Percent Identity: 42.169
XX
XX alignment_block:
XX US-08-973-363-11 x AAK06973/rev ..
XX
XX Align seg 1/1 to reverse of: AAK06973 from: 1 to: 475
XX
XX 7 LYSHTLSLEuHLSLysLSLleLysThrcLULysGLuSngLUGLULysPr 23
XX |||::: ||::| |:: |||||::| |||||
XX 364 AACAAGAGAAAGAGACAGACAGAGAGAGAGAGAGAGAGAGAGCA 315
XX 23 ocLUPrOAsrLleGLYLleLysLysGLuAlaGLUGLULysArgGLuThrL 40
XX || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 314 AAGAGGAG.....AAAGAGGAGGAGAGAGAGAGAGAGAGAGAGCA 277
XX 40 YSGULSLySGLUaSLnLysArGLUeULySArGLULysLysGLULysGLU 56
XX |||::| | | | | | | | | | | | | | | | | | | | | | | | | |
XX 226 AAGAAGAGAGAGAGAGACAGACAGAGAGAGAGAGAGAGAGAGAGCA 177
XX 57 AsPLySLySGLUleULySGLUlys...AsPAsnLysGLULysArGLUas 72
XX ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 226 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 177
XX 72 nLysValLysGLUsErThrGLnLysGLULysGLUValLysGLUGLULys 88
XX : | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 176 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 128

```

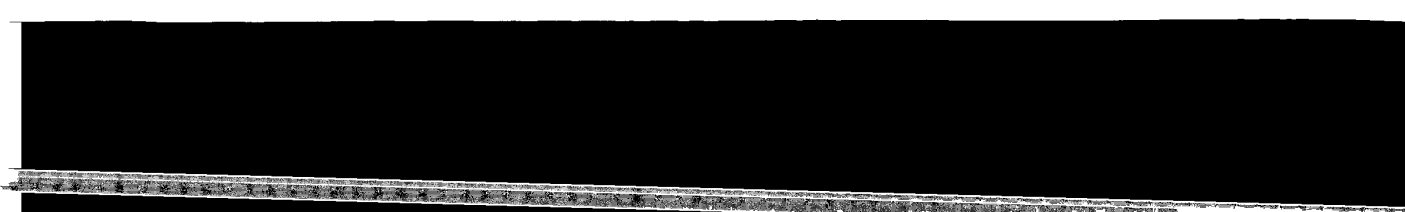




Mon Aug 5 11:51:39 2002

us-08-973-363-11.p2n.rng

Page 11



**THIS PAGE BLANK (USPTO)**







```

BASE COUNT      111 a      89 c      50 g      180 t      constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

alignment_scores:
    Quality: 271.00      Length: 68
    Ratio: 4.371      Gaps: 0
    Percent Similarity: 91.176      Percent Identity: 77.941

alignment_block:
US-08-973-363-11 x A1242163/rev ..

Align seg 1/1 to reverse of: A1242163 from: 1 to: 430

1  ASGCUlIleValSerValLysHISleuHISLysLysLleLysThrGluLys 17
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
206 GATGACATCAGTCTGTGTAACATCCAAATATAAAATTAACAGCAAG 157
17  SGLUASngLUGLUGLysProGluProAspLleGlyLleLysLysGluAlaG 34
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
156 AGACAGTGAAGAAACCTGACGACGATGTTATATTAAGAAAGAACG 107
34  LUGLUGLysArgGluThrLysGluLysGluLysLysArgGluLeuLysArg 50
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
106 AACAAAGAGAGCAAGCAAAAGAAAGCAATATAAAAGCACTTAAAG 57
51  GLUGLysLysGluLysGluAspLysLysGluLeuLysGluLysAspAsnLys 67
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
56 GAGATTAAGAAAGAAAGAGATTAAGCAATTAAGCAAAAGATTTTAA 7
67  SGLU 68
|||||
6  AGAA 3

seq_name: gb_est1:AA699918
seq_documentation_block:
LOCUS      AA699918      456 bp      mRNA      linear      EST 19-DEC-1997
DEFINITION  Z161f12.81 Soares_fetal_liver_spleen_lnf1s_S1 Homo sapiens cDNA
Clon IMAGE:435311 3', mRNA sequence.
ACCESSION  AA699918
VERSION    AA699918.1  GI:2702881
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 456)
AUTHORS   Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,
J., Moore, B., Schellenberg, K., Stepcoe, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 420.
Location/Qualifiers
1..456
/organism="Homo sapiens"
/db_xref="GDB:1335080"
/db_xref="taxon:9606"
/clone="IMAGE:435311"
/clone_lib="Soares_fetal_liver_spleen_lnf1s_S1"
/sex="male"

/dev stage="20 week-post conception fetus"
/lab host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site: 1: Pac I; Site 2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen lnf1s library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
ACTGCAGCAATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      119 a      95 c      57 g      185 t

ORIGIN

alignment_scores:
    Quality: 266.00      Length: 67
    Ratio: 4.361      Gaps: 0
    Percent Similarity: 91.045      Percent Identity: 77.612

alignment_block:
US-08-973-363-11 x AA699918/rev ..

Align seg 1/1 to reverse of: AA699918 from: 1 to: 456

1  ASGCUlIleValSerValLysHISleuHISLysLysLleLysThrGluLys 17
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
202 GATGACATCAGTCTGTGTAACATCCAAATATAAAATTAACAGCAAG 153
17  SGLUASngLUGLUGLysProGluProAspLleGlyLleLysLysGluAlaG 34
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
152 AGACAGTGAAGAAACCTGACGACGATGTTATTAAGAAAGAACG 103
34  LUGLUGLysArgGluThrLysGluLysGluLysLysArgGluLeuLysArg 50
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
102 AACAAAGAGAGCAAGCAAAAGCAATATAAAAGCACTTAAAG 53
51  GLUGLysLysGluLysGluAspLysLysGluLeuLysGluLysAspAsnLys 67
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
52 GAGATTAAGAAAGAGATTAAGCAATTAAGCAAAAGATTTTAA 3
67 s 67
|
2 A 2

seq_name: gb_est2:BF239967
seq_documentation_block:
LOCUS      BF239967      821 bp      mRNA      linear      EST 14-NOV-2000
DEFINITION  601905170F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:413129 5',
mRNA sequence.
ACCESSION  BF239967
VERSION    BF239967.1  GI:11153890
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 821)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgs@bbs-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

```

Plate: LCM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1. 821

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4133129"  
 /clone\_lib="NH\_MGC-54"  
 /tissue\_type="from chronic myelogenous leukemia"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: Still (ggccgcctcgcc); Site\_2: Still (ggccatctggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor sequence: 5'-ATTCGAGGCGCGCGCGCGCATG-GT(30)BN-3' (where B = A, C or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN

alignment\_scores:  
 Quality: 266.00 Length: 90  
 Ratio: 3.367 Gaps: 2  
 Percent Similarity: 87.778 Percent Identity: 68.889

alignment\_block:

US-08-973-363-11 x BF239967 ..

Align seg 1/1 to: BF239967 from: 1 to: 821

```

1  AspgllleValSerVallyshisleuHisLysLysLleLysThGluLy 17
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 GATGAGATCAGTCTGTGTAACATCCAAATTAATAAAACAGAAAG 283
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
17 sglu.AsnGluLysProGluProAspIleGlyLleLysLysGluAla 33
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
284 AGACCAAGTGAAGAAAACCTGAGCCAGATGTTTATATTAAGAACGACCA 333
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
34 GluGluLysArgGluThrLysGluLysGluAsnLysArgGluLeuLysAr 50
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 GAAGAAAGAGGAGCAAAAGCAAGATTAATAAAAGCACTTAAGAAAG 383
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
50 g.GluLysLysGluLysGluAspLysLysGluLeuLysLysAspAsn 66
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
384 GTGAGATTAAGAAAAGAGGATAGAAAGATAT.AAGGACACAGATTTT 432
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 LysGluLysArgGluAsnLysValLysGluLysGluSerThrGluLysGluLysGlu 83
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
433 AAAGAAAACGAGAAACCAAGTAAGAACCTTATACAGAAAGAAAG 482
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
83 uVallysGluGluLys 88
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
483 ACTAAAGCAGCAAAAG 498

```

seq\_name: gb\_est2:BM168938

seq\_documentation\_block:

LOCUS BM168938 660 bp mRNA linear EST 04-DEC-2001  
 DEFINITION EST571461 PYBS Plasmodium yoelii yoelii cDNA clone pYCP56 5' end,  
 mRNA sequence.

ACCESSION BM168938 GI:17302170

VERSION BM168938.1

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.

ORGANISM Plasmodium yoelii yoelii.

REFERENCE 1 (bases 1 to 660)

AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,T.W., Valdiva,A.B.,

TITLE Fraser,C.M. and Carucci,D.J.  
 JOURNAL Plasmodium yoelii EST project at TIGR  
 COMMENT Unpublished (2001)

CONTACT Contact: Jane Carlton  
 Parasite Genomics Group  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-530-9319  
 Fax: 301-838-0208  
 Email: carlton@tigr.org  
 For clone info, please contact the Malaria Research and Reference  
 Reagent Resource Center, ATCC  
 http://www.malaria.mr4.org/mr4pages/index.html  
 Seq primer: ADF.

FEATURES  
 source  
 Location/Qualifiers  
 1. 660

/organism="Plasmodium yoelii yoelii"  
 /strain="17XL"  
 /db\_xref="taxon:73239"  
 /clone="PYCPJ56"  
 /clone\_lib="PYBS"  
 /dev\_stage="Asexual blood stages"  
 /lab\_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size fractionated cDNA was precipitated and ligated to Hydrizap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the Hydrizap vector and plasmid DNA isolated."

BASE COUNT 367 a 35 c 143 g 115 t  
 ORIGIN

alignment\_scores:  
 Quality: 195.00 Length: 78  
 Ratio: 3.047 Gaps: 0  
 Percent Similarity: 82.051 Percent Identity: 48.718

alignment\_block:

US-08-973-363-11 x BM168938 ..

Align seg 1/1 to: BM168938 from: 1 to: 660

```

11 LysLysLleLysThGluLysGluAsnGluLysProGluProAspIle 27
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
4 CCCGAGTTGAATAAGTGAAGAAAGCAAAAGCAAAAGCAAAAGAA 53
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
27 eGlyLleLysGluAlaGluLysArgGluThrLysGluLysGluAla 44
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
54 AGAAGAAAAGAGATGAAGAAAGCAAGAAAGCAAGAAAGAAAGAA 103
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
44 snLysArgGluLeuLysArgGluLysGluLysGluAspLysGlu 60
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
104 AAAAGAAATTAAGCAATTAAGAAAGCAAGCAAGCAATTAAGCA 153
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
61 LeuLysGluLysAspAsnLysGluLysArgGluAsnLysValLysGlu 77
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
154 GAAAGGAAGAGATTAAGAAAGCAAGAAAGCAAGATTAAGCAAG 203
   ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
77 rThrGluLysGluLysGluValLysGluLys 88
   ::::: ||||||| ||||||| ||||||| ||||||| |||||||
204 AAAAGAAAGAAAGAAAGCAAGAAAGCAAGAAAG 237

```

```

seq_name: gb_est2:BM161742
seq_documentation_block:
LOCUS      BM161742                605 bp      mRNA      linear      EST 04-DEC-2001
DEFINITION      EN556265 PYNS Plasmodium yoelii yoelii cDNA clone PYCK149 5' end,
ACCESSION      BM161742
VERSION        BM161742
KEYWORDS       EST
SOURCE         Plasmodium yoelii yoelii.
ORGANISM       Plasmodium yoelii yoelii.
REFERENCE      1 (bases 1 to 605)
AUTHORS        Fraser,C.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,
              Carlson,J.M., and Carucci,D.J.
TITLE          Plasmodium yoelii EST project at TIGR
JOURNAL        Unpublished (2001)
COMMENT        Contact: Jane Carlson
                Parasite Genomics Group
                The Institute for Genomic Research
                9712 Medical Center Drive, Rockville, MD 20850, USA
                Tel: 301-530-9319
                Fax: 301-838-0208
                Email: carlont@igf.org
                For clone info, please contact the Malaria Research and Reference
                Reagent Resource Center, ATCC
                http://www.malaria.mr4.org/mr4pages/index.html
                Seq primer: ADP.
FEATURES
     source          Location/Qualifiers
                     1..605
                     /organism="Plasmodium yoelii yoelii"
                     /strain="17XL"
                     /db_xref="taxon:73239"
                     /clone="PYCK149"
                     /clone_1lb="PYNS"
                     /dev_stage="Asexual blood stages"
                     /lab_host="E. coli XL-1 Blue"
                     /note="Vector: pAD-GAL4. At 20-25% parasitemia, blood was
                     collected from BALB/cByJ mice infected with Py17XL
                     parasites, and leukocytes removed by passage over
                     microcrystalline cellulose columns. Total RNA was
                     isolated using the guanidium thiocyanate method, and
                     mRNA isolated using oligo(dT)-cellulose chromatography.
                     First strand cDNA synthesis was completed using a 50-base
                     primer and reverse transcriptase in the presence of
                     5-methyl dCTP. After second strand synthesis, uneven
                     termini were treated with Pfu DNA polymerase and EcoRI
                     adaptors ligated to the blunt ends. The sample was cleaved
                     with XhoI and separated on a Sephacryl S-500 column.
                     Size-fractionated cDNA was precipitated and ligated to
                     HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
                     After packaging, the phagemid vector (pAD-GAL4) was
                     excised from the HybriZAP vector and plasmid DNA
                     isolated."
BASE COUNT      342 a          138 g          88 t
ORIGIN
alignment_scores:
      Quality: 193.00      Length: 77
      Ratio: 3.063      Gaps: 0
Percent Similarity: 81.818      Percent Identity: 49.351
alignment_block:
US-08-973-363-11 x BM161742      ..
Align seg 1/1 to: BM161742 from: 1 to: 605
12 LysIIlelysthrGlysgtLuanngLuglulysprGcugluprpsIleel 28
      .....|||||.....
8 GAAGTGAAGAAAGCTGAGAAAGTGAAGAAAGAAAGAAAGAAAGAA 57

```

```

28 ytttlytsgtltatagltgttlytsgtltgtthrttsgtltgtatant 45
|||||: : : : : |||||: : : : :
58 AGAAAAAAGCATGAAAGCAAGAAAGCAAGAAAGCAAGAAAGCAAGAA 107
|||||: : : : : |||||: : : : :
45 ttttsgtltgtltatgtgttlytsgtltgtatgttlytsgtltgt 61
|||||: : : : : |||||: : : : :
108 AAGAGATATAGCAAGATATATATATATATATATATATATATATAT 157
|||||: : : : : |||||: : : : :
62 ttttsgtltgtatgttlytsgtltgtatgttlytsgtltgtatgt 78
|||||: : : : : |||||: : : : :
158 AAGAGAGATATATATATATATATATATATATATATATATATAT 207
|||||: : : : : |||||: : : : :
78 ttttsgtltgtatgttlytsgtltgtatgttlytsgtltgt 88
|||||: : : : : |||||: : : : :
208 AAGAGAGATATATATATATATATATATATATATATATATAT 238
|||||: : : : : |||||: : : : :

seq_name: gp_est2:BM167712

seq_documentation_block:
LOCUS BM167712 640 bp mRNA linear EST 04-DEC-2001
DEFINITION EST570235 PyBS Plasmidium yoeIII yoeIII cDNA clone pYCOP51 5' end,
MRNA sequence.
ACCESSION BM167712
VERSION BM167712.1 GI:17300944
KEYWORDS EST.
SOURCE Plasmidium yoeIII yoeIII.
ORGANISM Plasmidium yoeIII yoeIII.
REFERENCE 1 (bases 1 to 640)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valday,A.B.,
Eukaryota, Alveolata, Apicomplexa, Haemosporida, Plasmidium.
TITLE Plasmidium yoeIII yoeIII.
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@igf.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center. ARCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES
            Location/Qualifiers
             ..640
             /organism="Plasmidium yoeIII yoeIII"
             /strain="17XL"
             /db_xref="taxon:73239"
             /clone="pYCOP51"
             /clone_1b="PyBS"
             /dev_stage="Asexual blood stages"
             /lab_host="E. coli XL-1 Blue"
             /note="Vector: pBD-GAL4: At 20-25% parasitemia, blood was
             collected from BALB/cBYJ mice infected with Pyl7XL
             parasites, and leukocytes removed by passage over
             microcrystalline cellulose columns. Total RNA was
             isolated using the guanidinium isothiocyanate method, and
             first strand cDNA synthesis was completed using a 50-base
             primer and reverse transcriptase in the presence of
             5methyl dCTP. After second strand synthesis, uneven
             termini were treated with Pfu DNA polymerase and EcoRI
             adaptors ligated to the blunt ends. The sample was cleaved
             with XhoI and separated on a Sephacryl S-500 column.
             Size-fractionated cDNA was precipitated and ligated to
             HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
             After packaging, the phagemid vector (pAD-XhoI) was
             excised from the HybriZAP vector and plasmid DNA
             isolated."
BASE COUNT      358 a      34 c      144 g      104 t
ORIGIN

```



```

/organism="Sturana tropicalis"
/db_xref="taxon:8364"
/clone="TCG8009e08"
/dev_stage="stg gastrula"
/lab_host="Escherichia coli XL1-blue"
/label="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5' end of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT      263 a      64 c      122 g      56 t      2 others
ORIGIN

alignment_scores:
  Quality:      189.50      Length:      85
  Ratio:        2.828      Gaps:        2
  Percent Similarity: 78.824      Percent Identity: 44.706

alignment_block:
US-08-973-363-11 x AL628267 ..

Align seg 1/1 to: AL628267 from: 1 to: 507

1 AspgtllleValSerValLysHsLeuHsLysLysLysLysThgGly 17
:::||||| ::::: ::::: ::::: ::::: :::::
256 GAGGAAACTGTACCCCTCAGACAAATATAATAAG...AGAGCAGAGAA 302
17 sgluasnglulysprogluproaspriegllytleLysGluAlaG 34
||||| ::::: ::::: ::::: ::::: :::::
303 AGAAATGACGACGACGACGACGACGACGACGACGACGACGACG 352
34 lugalysarglulthrysglulysgluasnllysarglulysarg 50
||||| ::::: ::::: ::::: ::::: :::::
353 AAGAAAAGAAAGAGATA....AAGAAATATAAAGACCCCTAAGAA 396
51 glulyslsglulysgluasplyslsglulysglulysaspsanly 67
||||| ::::: ::::: ::::: ::::: :::::
397 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 446
67 sgluasnglulysarglulthrysglulysglulysglulysg 84
::: ::::: ::::: ::::: ::::: :::::
447 AGACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 496
84 allys 85
|||
497 AAAAA 501

seq_name: gb_est1:A1890775
seq_documentation_block:
LOCUS      A1890775      547 bp      mRNA      linear      EST 07-MAR-2000
DEFINITION      wmsjfl1.x1 NCI-CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'
similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
PROTEIN 1; mRNA sequence.
ACCESSION      A1890775      GI:5595939
VERSION      A1890775.1
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 547)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL      Tumor Gene Index
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA library Preparation: Life Technologies, Inc.

```

```

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLM at:
www.bio.linn.gov/biopr/image/image.html
Insert Length: 1924      Std Error: 0.00
Seq primer: -400P from Glbco
High quality sequence stop: 418.
Location/Qualifiers
1. 547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2443725"
/clone_lib="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
BASE COUNT      114 a      118 c      85 g      230 t
ORIGIN

alignment_scores:
  Quality:      175.00      Length:      44
  Ratio:        4.375      Gaps:        0
  Percent Similarity: 90.909      Percent Identity: 77.273

alignment_block:
US-08-973-363-11 x A1890775/rev ..

Align seg 1/1 to reverse of: A1890775 from: 1 to: 547

1 AspgtllleValSerValLysHsLeuHsLysLysLysLysThgGly 17
||||| ::::: ::::: ::::: ::::: :::::
132 CATAGATCATGCTCTGTAAACATCAATATAAATATAAAGAGAG 83
17 sgluasnglulysprogluproaspriegllytleLysGluAlaG 34
::: ::::: ::::: ::::: ::::: :::::
82 AGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33
34 lugalysarglulthrysglulysgluasn 44
||||| ::::: ::::: ::::: ::::: :::::
32 AAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1

seq_name: gb_gss:A2008814
seq_documentation_block:
LOCUS      A2008814      634 bp      DNA      linear      GSS 25-FEB-2000
DEFINITION      RPCI-23-38903.TJ RPCI-23 Mus musculus genomic clone RPCI-23-38903,
DNA sequence.
ACCESSION      A2008814
VERSION      A2008814.1
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 634)
AUTHORS      Zhao,S., Nierman,W., Feldblum,T., Malek,D., Shatsman,S., Akimret,
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLES      Mouse BAC End Sequences from Library RPCI-23
JOURNAL      Unpublished (1999)
COMMENT      Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

```

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Reseach Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)

Plate: 389 row: 0 column: 3  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source location/Qualifiers

1. 634  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-38903"  
/clone\_id="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 278 a 73 c 219 g 63 t 1 others  
ORIGIN

alignment\_scores:

Quality: 173.00 Length: 95  
Ratio: 2.703 Gaps: 4  
Percent Similarity: 67.368 Percent Identity: 47.368

alignment\_block:

US-08-973-363-11 x AZ008814 ..

Align seg 1/1 to: AZ008814 from: 1 to: 634

```
7 LysH1sleuH1s.....Lys1s1le1y5ThrGlu1y 17
||||| ||| ||||| ||| |||||
61 AAGCATCATCACCAGTCTTGTATGAGAAAGAAAGAGAGAGAGAGA 110
17 sgluansnglu1y1y5ProgluProasp1leGly1le1y5glu... 32
||||| ||||| ||||| ||||| |||||
111 GGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 160
33 .....AlaGluGlu1y5ArgGluThr1y5Glu1y5glu1y5 45
||||| ||||| ||||| ||||| |||||
161 AGAAGGAGATTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 210
46 ArgGlu1y5ArgGlu1y5Lys1y5glu1y5glu1y5Lys1y5glu1y5 62
||||| ||||| ||||| ||||| |||||
211 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 257
62 sglu1y5spasn1y5glu1y5ArgGlu1y5Val1y5glu1y5Thr1y5 79
||||| ||||| ||||| ||||| |||||
258 GAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 307
79 lnt1y5glu1y5.....GluVal1y5glu1y5 88
||||| ||||| ||||| ||||| |||||
308 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
```

seq\_name: gb\_gss:AZ330985

seq\_documentation\_block:

LOCUS AZ330985 743 bp DNA linear GSS 29-SEP-2000  
DEFINITION IM0056C22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
ACCESSION AZ330985  
VERSION AZ330985.1 GI:10393428  
KEYWORDS GSS.  
SOURCE house mouse.

#### ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 743)

#### REFERENCE

1 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly  
,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhauser,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

#### TITLE

#### JOURNAL

#### COMMENT

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0056 row: C column: 22  
Seq primer: CACACAGAGAAACGATATGACC  
Class: plasmid ends  
High quality sequence stop: 743.

FEATURES  
source

Location/Qualifiers

1. 743

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0056C22"

/clone\_id="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adapter oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PMD42 (g14732114[gb|AF129072.1]), a copy-number  
inducible derivative of plasmid RL. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 107 a 256 c 102 g 278 t

ORIGIN

alignment\_scores:

Quality: 172.00 Length: 78  
Ratio: 2.730 Gaps: 3  
Percent Similarity: 80.769 Percent Identity: 51.282

alignment\_block:

US-08-973-363-11 x AZ330985/rev ..

Align seg 1/1 to reverse of: AZ330985 from: 1 to: 743

```
11 Lys1y51le1y5ThrGlu1y5nglu1y5ProgluProasp1l 27
||||| ||| ||||| ||||| ||||| |||||
743 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 699
27 eGly1le1y5glu1y5AlaGlu1y5ArgGlu1y5Thr1y5glu1y5glu 44
||||| ||||| ||||| ||||| ||||| |||||
698 .....AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656
```



## COMMENT

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

## source

1. .981  
/organism="Debaryomyces hansenii"  
/strain="CBS 767"  
/variety="hansenii"  
/db\_xref="taxon:4959"  
/clone="BC0A009H03"  
/clone\_1lb="BC0A"  
/note="end : T3"  
<211. .>665

## misc\_feature

/note="similar to *Saccharomyces cerevisiae* ORF YGR280C [ weak similarity to Cbf5p ]  
1 putative frameshift(s)"  
/evidence="not\_experimental"

BASE COUNT 430 a 98 c 241 g 209 t 3 others  
ORIGIN

## alignment\_scores:

Quality: 170.00 Length: 94  
Ratio: 2.656 Gaps: 4  
Percent Similarity: 68.085 Percent Identity: 45.745

## alignment\_block:

US-08-973-363-11 x CNS07B7L ..

Align seg 1/1 to: CNS07B7L from: 1 to: 981

```

5  SerValTysHIsLseuHIsLysLysIle..... 13
   ||| :||| |||||
   ||| :||| |||||
711 AGTAGCAGCATTAAGAGAAAAATAAAGAGAGACCGTAAGAAAAAGA 760
   ||| :||| |||||
14  .....LysThrGluLysGluAsnGluGluLysProGluProAspIleG 28
   ||||| ||||| ||||| |||||
761 AGAGAGAGAGACCGAGAGAAATATGTGAG..... 791
   ||| :||| |||||
28  LylLeLysLysGluLacGluLysArgGluThrLysGluLysGluAsn 44
   ||||| ||||| :||| ||||| ||||| |||||
792 ..ATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
   ||| :||| ||||| ||||| ||||| |||||
45  LysArgGluLeuLysArgGluLysLysGluLysGluAspLysLysGlu 61
   ||| :||| ||||| ||||| ||||| ||||| |||||
840 AAGARAGAT...AAGACGAGAGAGAGAGAGAGAGATTAAGACGAGAA 886
   ||| :||| ||||| ||||| ||||| ||||| |||||
61  uLysGluLysAspAsnLysGluLysArgGluAsnLysValLysGluSer 78
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
887 GAAAGAGAGAG.....AAGATTAAGACGAGAGAGAGAGAGAGAGAA 930
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
78  hrGluLysGluLysGluValLysGluGluLys 88
   :||| :||| ||||| ||||| ||||| ||||| |||||
931 AGAAGAGAGAGAGAGAGAGAGATTAAGAAA 962

```







```

REFERENCE 2 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
3 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.
Draft Sequence Produced by DOE Joint Genome Institute
WWW: jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
WWW: shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 146.7kb). It is clipped at the overlap with AC012624.
The number of bases overlapped is 90404.
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-58M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN
alignment_scores:
Quality: 159.00 Length: 45
Ratio: 4.077 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 68.889
alignment_block:
US-08-973-363-14 x AC092372/rev ..
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220
1 Asgcgyllevalserallyshisprohislyslileysalagluly 17
|||||
25410 GATGAGATCACTTCGTGAAACATCCAAATTTAAACACAGAAAG 25361
17 sgluysnglululysaspcluprogliuilegyllelyslugluag 34
|||||
25360 AGCAGGTGAAAGAAAACCTGACCCAGATGTTATTTAAAGACGAACG 25311
34 lyglulysarglurhrlysluglulysgluasnly 45
|||||
25310 AAGAAAAGAGGAAAGCAAAAGAAAGAGAGATTA 25276
seq_name: gb_pr:AC012624
seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-208211, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM human.
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

```

REFERENCE      3 (bases 1 to 134365)
AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Direct Submission
JOURNAL       Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
REFERENCE     4 (bases 1 to 134365)
AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Direct Submission
JOURNAL       Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT       On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
   source
       1..134365
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="5"
           /clone="CTD-2082117"
BASE COUNT    40414 a 24497 c 25503 g 43951 t
ORIGIN
alignment_scores:
    Quality: 159.00          Length: 45
    Ratio: 4.077            Gaps: 0
Percent Similarity: 86.667 Percent Identity: 68.889
Alignment_block:
US-08-973-363-14 x ACO12624 ..
Align seg 1/1 to: ACO12624 from: 1 to: 134365
1 ASGGCTTTCAGATGACGTTCTGTGAACCATCCAAATTAATAAAGCGAAG 119773
119773 GATGAGATCAGTTCTGTGAACCATCCAAATTAATAAAGCGAAG 119822
17 GCGAACGCTGATCAAAAACCTGCACCAGATTGTTATTAAGAAGACACGAG 119823
119823 AGCATGTAAGAAAACTGCACCAGATTGTTATTAAGAAGACACGAG 119872
34 TCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 119873
seq_name: gb_htg:ACO21449
seq_documentation_block:
LOCUS        ACO21449                143079 bp    DNA             linear      HTG 10-SEP-2000
DEFINITION   Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION    ACO21449
VERSION      ACO21449.3 GI:10047806
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 143079)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 143079)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,P., Bedalov,F.,
Bonuslinsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Chapelat,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRubeis,K., Demet,K., Domini,M., Doyle,M., Fensholt,J.,
Fitzell,R., Fitzhugh,M., Forrest,C., Gage,D., Galagan,J.,
Gerard,J.S., Grant,G., Hagos,B., Heald,A., Horton,L.,
Kandathil,J.C., Johnson,S., Jones,C., Kamm,L., Karasik,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McLean,P., McGuck,A., McKernan,K.,
McNeehters,R., Melidze,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,

```









```
* 27984 30386: contig of 2403 bp in length
* 30387 30486: gap of unknown length
* 30487 32480: contig of 1994 bp in length
* 32481 32580: gap of unknown length
* 32581 35441: contig of 2861 bp in length
* 35442 35541: gap of unknown length
* 35541 37801: contig of 2259 bp in length
* 37801 37901: gap of unknown length
* 37901 40587: contig of 2687 bp in length
* 40588 40687: gap of unknown length
* 40688 44159: contig of 3472 bp in length
* 44160 44259: gap of unknown length
* 44260 46636: contig of 2377 bp in length
* 46637 46736: gap of unknown length
* 46737 50082: contig of 3346 bp in length
* 50083 50183: gap of unknown length
* 50183 53989: contig of 3806 bp in length
* 53989 54088: gap of unknown length
* 54089 56592: contig of 2504 bp in length
* 56593 56692: gap of unknown length
* 56693 61352: contig of 4660 bp in length
* 61353 61452: gap of unknown length
* 61453 64254: contig of 2802 bp in length
* 64255 64354: gap of unknown length
* 64355 68825: contig of 4471 bp in length
* 68826 68925: gap of unknown length
* 68926 74393: contig of 5468 bp in length
* 74394 74493: gap of unknown length
* 74494 80561: contig of 6068 bp in length
* 80562 80661: gap of unknown length
* 80662 87626: contig of 6965 bp in length
* 87627 87727: gap of unknown length
* 87727 93601: contig of 5874 bp in length
* 93601 93701: gap of unknown length
* 93701 102024: contig of 8324 bp in length
* 102025 102124: gap of unknown length
* 102125 109905: contig of 7781 bp in length
* 109906 110005: gap of unknown length
* 110006 116737: contig of 6732 bp in length
* 116738 116837: gap of unknown length
* 116838 126797: contig of 9960 bp in length
* 126798 126897: gap of unknown length
* 126898 135114: contig of 8217 bp in length
* 135115 135214: gap of unknown length
* 135215 146713: contig of 11499 bp in length
* 146714 146813: gap of unknown length
* 146814 157894: contig of 11081 bp in length
* 157895 157994: gap of unknown length
* 157995 171752: contig of 13758 bp in length
* 171753 171852: gap of unknown length
* 171853 201687: contig of 29835 bp in length
* 201688 231268: gap of unknown length
* 231269 231368: gap of unknown length
* 231369 276181: contig of 44813 bp in length.
```

## FEATURES

source

```
1. 276181
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="5"
   /clone_lib="RP11-75H1"
   /clone_1lb="RP11 human BAC library 11"
BASE COUNT      86859 a 51769 c 50026 g 82854 t 4673 others
ORIGIN
```

```
alignment_scores:
  Quality: 159.00      Length: 45
  Ratio: 4.077         Gaps: 0
Percent Similarity: 86.667 Percent Identity: 68.889
```

```
alignment_block:
US-08-973-363-14 x AC092382/rev ..
```

```
Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181
1 AspglyIleValSerValIysHisProHisIysIleIysAlaGluIly 17
|||||
212122 GATGAGATCAGTCTCTGTGTAACATCCATAATAAATTAAACAGAAAG 212123
17 sgluAnsgluGluIysAspgluProGluIleGlyIleIysIysGluAlaG 34
212122 AGACAGTCAGAAAAAAGCTGAGCCAGATGTTATATTAAGAAGAACAG 212073
34 IyGluIysArgGluThrIysGluIysGluAsnIys 45
212072 AAGAAAGAGGAGCAAGCAAAAGAAAGAGAAATATAA 212038
```

seq\_name: gb\_pr:HS1185H19

seq\_documentation\_block:

LOCUS

DEFINITION

Human DNA sequence from clone RP5-1185H19 on chromosome 1p13.1-13.3  
Contains part of a gene for a novel protein, STSS, GSSS and a Cpg  
island, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

```
1. 104780
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="1"
   /map="p13.1-13.3"
   /clone="RP5-1185H19"
   /clone_1lb="RP1-5"
   repeat_region
   /note="LIM4 repeat: matches 2452..3097 of consensus"
```





alignment\_scores:                   Quality:           91.00                   Length:           38  
                                     Ratio:               3.138                   Gaps:             0  
Percent Similarity: 76.316       Percent Identity: 50.000

alignment\_block:

US-08-973-363-14 x HS1185H19 ..

Align seg 1/1 to: HS1185H19 from: 1 to: 104780

```
8 H18P8H185H19LysIleLysAlaGluLysGluGluGluLysAspG1 24
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
97423 TATCCAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 97472
24 uP8G1u185H19LysIleLysAlaGluLysGluGluGluLysG 41
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
97473 GGAAGGAAAGAAATAAAGAAAGAAAGAAAGAAAGAAAGAAAG 97522
```

41 LulysGluAsnLys 45  
|||||:::|||||  
97523 AAAGAAAAAGAAA 97536

seq\_name: gb\_ro:AC067964

seq\_documentation\_block:  
LOCUS AC067964 219769 bp DNA linear ROD 05-APR-2001  
DEFINITION Mus Musculus strain C57BL6/J Chromosome 5 RP23-337K7, complete  
sequence.  
ACCESSION AC067964  
VERSION AC067964.18 GI:13549252  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,  
1 (bases 1 to 219769)  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
High Throughput Mouse Sequencing  
Unpublished  
2 (bases 1 to 219769)  
Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
Direct Submission  
Submitted (27-APR-2000) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA  
3 (bases 1 to 219769)  
Han,J., Montgomery,K.T., Grills,G., Lee,E., Long,J., Pomerantz,R.,  
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,  
Gordon,M., Goltz,J.S. and Kucherlapati,R.  
Direct Submission  
Submitted (05-APR-2001) Department of Molecular Genetics, Albert  
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,  
Bronx, NY 10461, USA

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

COMMENT  
-----Genome Center:  
Harvard Partners Genome Center  
Center Code: HPGC  
Web site:  
http://wchanning.bwh.harvard.edu:9088/hpcpg/jsp/hpcpg/sequence/mous  
e.html  
Contact: gntm@capcod.bwh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this  
clone unless otherwise noted. If there are overlapping clones, the  
overlaps are noted in the beginning and end of the Features  
listing.

ANNOTATION OF FEATURES:  
STS are identified using ePCR (Genome Res. 7:541-550).

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Regions of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST  
and cDNA sequences in Unigene. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintain sequence continuity  
across the splice junctions. Sequences that are not identical  
matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double  
stranded sequence for all regions. All sequence is completed to a  
standard of coverage with a minimum of 3 reads with no ambiguities.  
If the sequence coverage for a region does not meet this standard,  
it is indicated in the annotation as Low Coverage. Low coverage  
linkages are verified by PCR product size verification or  
verification of forward and reverse reads from clones which span  
the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated average error rate is less than 1 per 10,000  
bases using the Consed quality parameters. Regions that do not  
meet this requirement are annotated as Low Quality.

-----Summary Statistics  
Center project name: ACS  
Sequencing vector: pUC18; 108752  
Chemistry: Dye-terminator Big Dye; 100#  
Assembly program: Phrap version 0.990319  
Contig length: 219769  
Fraction of Phrap value < 40: 0.00293  
Error Rate in Consed: 0.07 per 10,000 bases  
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

	10001	9001	8001	7001	6001	5001	4001	3001	2001	1001	01
#	10001	9001	8001	7001	6001	5001	4001	3001	2001	1001	01
bases	10001	9001	8001	7001	6001	5001	4001	3001	2001	1001	01

-----  
Phrap Value Range

FEATURES  
-----  
source  
1. Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/clone="RP23-337K7"  
1. 108900  
/note="Overlap with Adjacent Clone RP23-301H20, AC077689"  
complement(268..359)  
/rpt\_family="LIM1"  
repeat\_region  
627..703  
/rpt\_family="(TCMTCG)n"  
704..757  
/rpt\_family="CT-rich"  
719..761  
/rpt\_family="(TCG)n"  
complement(773..867)  
/rpt\_family="PBID9"  
complement(913..1027)  
/rpt\_family="LIM1"  
repeat\_region  
1206..1406  
/rpt\_family="B3"  
repeat\_region  
1420..1440

repeat_L_region	/rpt_family="AT_rich"	complement(1491. .1637)
repeat_L_region	/rpt_family="B1_MW"	1648. .1688
repeat_L_region	/rpt_family="AT_rich"	1786. .1822
repeat_L_region	complement("TTTTTC)n"	complement(11828. .1958)
repeat_L_region	/rpt_family="B1_MW"	complement(12006 .2367)
repeat_L_region	/rpt_family="ORR1C"	3600. .3772
repeat_L_region	/rpt_family="B3"	3773. .3833
repeat_L_region	/rpt_family="CA)n"	3834. .3850
repeat_L_region	/rpt_family="B3"	complement(4270. .4469)
repeat_L_region	/rpt_family="B3A"	5341. .5370
repeat_L_region	/rpt_family="AT_rich"	5531. .5772
repeat_L_region	/rpt_family="MTE"	5774. .5803
repeat_L_region	/rpt_family="TTTTG)n"	complement(5829. .5845)
repeat_L_region	/rpt_family="B2_MW2"	5846. .5926
repeat_L_region	/rpt_family="TA)n"	complement(5927. .6073)
repeat_L_region	/rpt_family="B2_MW3"	6078. .6138
repeat_L_region	/rpt_family="MTE"	6443. .6622
repeat_L_region	/rpt_family="T1M1"	complement(7600. .8061)
repeat_L_region	/rpt_family="B1_MW"	complement(7603. .7618)
repeat_L_region	/rpt_family="B4A. .7931)	complement(7619. .7931)
repeat_L_region	/rpt_family="ORR1A3"	complement(7932. .8116)
repeat_L_region	/rpt_family="B4A"	8518. .8883
repeat_L_region	/rpt_family="CA)n"	9316. .9337
repeat_L_region	/rpt_family="AT_rich"	complement(9375. .9534)
repeat_L_region	/rpt_family="RSINE1"	9690. .9709
repeat_L_region	/rpt_family="CAAA)n"	9932. .9978
repeat_L_region	/rpt_family="CA)n"	10152. .10251
repeat_L_region	/rpt_family="CA)n"	10852. .10882
misc_feature	/standard_name="low coverage /note="GGA Repeat"	
repeat_L_region	/rpt_family="B3"	11222. .11273
repeat_L_region	/rpt_family="GGA)n"	11600. .11629
repeat_L_region	/rpt_family="G)n"	complement(11658. .11712)
repeat_L_region	/rpt_family="FBID"	13772. .13944
repeat_L_region	/rpt_family="RMR17C"	13945. .13993
repeat_L_region	/rpt_family="TC)n"	13994. .14145
repeat_L_region	/rpt_family="RMR17C"	complement(14702. .15092)
repeat_L_region	/rpt_family="MTB"	complement(15182. .15485)

```

repeat_region      /rpt_family="B7_Km1"
                    /rpt_family="B2_Km1"
repeat_region      complement(16697..16915)
repeat_region      /rpt_family="B3A"
repeat_region      17335..17378

Alignment_scores:
    Quality:      87.50          Length:      45
    Ratio:        2.734         Gaps:        1
    Percent Similarity: 71.111   Percent Identity: 44.444

Alignment_block:
US-08-973-363-14 x AC067964 ..

Align seg 1/1 to: AC067964 from: 1 to: 219769

1 AspglylIeValserIalyshtsIProHlsYslYslleYslsAlGuly 17
||| :::::::::::::::::::: |||||
201985 GATGTGGAAACAGCTTCAGTATGCATCGTAAGAAAAA.....GAAGA 202025

17 sgluasngluGluyAspGIuProGluIueGlyllyLyLysGluAlua 34
||||:|||||:::||| ||| ::|||::||| |
202026 TGAAGAGAGCAGCAGAGCAAGAAAAGCAGCAGGTCAAAGAGAGG 202075

34 lyGluYsArgGluThrLySgLuYsGluYslu 45
||||:|||||:::||| |||||:::||| |||
202076 GACAGAGAGAGCAAGAAAGCAAGAGGAAAGGAAAG 202110

seq_name: gb_to:AL589701

seg_documentation_block:
LOCUS       AL589701             219200 bp     DNA           linear     ROD 30-JAN-2002
DEFINITION  Mouse DNA sequence from clone RP23-202F3 on chromosome 13.
ACCESSION   AL589701
VERSION     AL589701.9  GI:18476660
KEYWORDS    HTG.
SOURCE      mouse mouse.
ORGANISM    Mus musculus.
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            I (bases 1 to 219200)
REFERENCE   1
AUTHORS    Phillimore,B.
TITLE      Direct Submission
JOURNAL    Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humgvery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On Feb 11, 2002 this sequence version replaced gi:117384104.
            During difference assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences without
            only a small overlap as described above.
COMMENT     This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one Plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em., EMBL, Sw.,
            SWISSPROT; Tr., TREMBL; Wp., WormPep; Information on the WormPep
            database can be found at
            http://www.sanger.ac.uk/projects/C_elegans/wormpep
            from the RPI-23 Mouse PAC Library
            constructed by the group of Pieret de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBlac3.6
            This sequence is the entire insert of clone RP23-202F3. The true
            left end of clone RP23-171015 is at 179556 in this sequence. The
            true right end of clone RP23-153B6 is at 117857 in this sequence.

```



```
* 71297 77926: contig of 6630 bp in length
* 77927 78026: gap of unknown length
* 78027 84372: contig of 6346 bp in length
* 84373 84473: gap of unknown length
* 84474 90426: contig of 5957 bp in length
* 90427 90430: gap of unknown length
* 90431 94563: gap of 4040 bp in length
* 94564 94570: gap of unknown length
* 94571 100480: contig of 5811 bp in length
* 100481 100580: gap of unknown length
* 100581 100586: contig of 5056 bp in length
* 100587 105737: gap of unknown length
* 105738 105739: contig of 4106 bp in length
* 105740 109943: gap of unknown length
* 109944 109945: contig of 4916 bp in length
* 109946 114859: gap of unknown length
* 114860 114859: contig of 3576 bp in length
* 114860 118535: gap of unknown length
* 118536 123663: contig of 5029 bp in length
* 123664 123763: gap of unknown length
* 123764 127748: contig of 3985 bp in length
* 127749 127849: gap of unknown length
* 127850 131812: contig of 3964 bp in length
* 131813 131912: gap of unknown length
* 131913 135499: contig of 3587 bp in length
* 135500 135599: gap of unknown length
* 135600 139874: contig of 4275 bp in length
* 139875 139974: gap of unknown length
* 139975 144298: contig of 4324 bp in length
* 144299 144398: gap of unknown length
* 144399 148526: contig of 4128 bp in length
* 148527 148626: gap of unknown length
* 148627 152730: contig of 4104 bp in length
* 152731 152830: gap of unknown length
* 152831 157579: contig of 4749 bp in length
* 157580 157679: gap of unknown length
* 157680 161266: contig of 3587 bp in length
* 161267 161366: gap of unknown length
* 161367 164887: gap of 3621 bp in length
* 164888 165087: gap of unknown length
* 165088 167888: contig of 2801 bp in length
* 167889 168088: gap of unknown length
* 168089 170166: contig of 2078 bp in length
* 170167 170266: gap of unknown length
* 170267 172944: contig of 2678 bp in length
* 172945 173044: gap of unknown length
* 173045 176120: contig of 3076 bp in length
* 176121 176220: gap of unknown length
* 176221 178990: contig of 2770 bp in length
* 178991 179090: gap of unknown length
* 179091 181701: contig of 2611 bp in length
* 181702 181801: gap of unknown length
* 181802 183811: contig of 2010 bp in length
* 183812 183911: gap of unknown length
* 183912 185953: contig of 2042 bp in length
* 185954 186053: gap of unknown length
* 186054 187891: contig of 1838 bp in length
* 187892 187991: gap of unknown length
* 187992 190795: contig of 2804 bp in length
* 190796 190895: gap of unknown length
* 190896 193028: contig of 2133 bp in length
* 193029 193128: gap of unknown length
* 193129 195041: contig of 1913 bp in length
* 195042 195141: gap of unknown length
* 195142 196166: contig of 1275 bp in length
* 196167 198846: gap of unknown length
* 198847 198946: gap of unknown length
* 198947 201855: contig of 2309 bp in length
* 201856 201955: gap of unknown length
* 201956 203450: contig of 1495 bp in length
* 203451 203550: gap of unknown length
* 203551 206062: contig of 2512 bp in length
```

```
* 206063 206162: gap of unknown length
* 206163 207721: contig of 1559 bp in length
* 207722 207821: gap of unknown length
* 207822 210574: contig of 2753 bp in length
* 210575 210674: gap of unknown length
* 210675 212188: contig of 1514 bp in length
* 212189 212288: gap of unknown length
* 212289 213701: contig of 1413 bp in length
* 213702 213801: gap of unknown length
* 213802 215447: contig of 1746 bp in length
* 215448 215647: gap of unknown length
* 215648 217053: contig of 1406 bp in length
* 217054 217153: gap of unknown length
* 217154 218235: contig of 1081 bp in length
* 218235 218334: gap of unknown length
* 218335 219342: contig of 1008 bp in length
* 219343 219442: gap of unknown length
* 219443 220550: contig of 1108 bp in length
* 220551 220650: gap of unknown length
* 220651 222211: contig of 1561 bp in length
* 222212 222311: gap of unknown length
* 222312 224092: contig of 1781 bp in length
* 224093 224192: gap of unknown length
```

```
alignment_scores:
  Quality: 87.00 Length: 43
  Ratio: 2.636 Gaps: 0
Percent Similarity: 76.744 Percent Identity: 39.535
```

alignment\_block:

US-08-973-363-14 x AC098600/rev ..

Align seg 1/1 to reverse of: AC098600 from: 1 to: 231947

```
3 11e1a1SeYAllyHs1ProH1sY1s1leYs1la1g1u1y1s1u1s 19
21458 ATGTATAATGTTAAACATAGCAAAAGAAAGAGAGAGAGAGAG 21409
19 nglug1u1y1s1a1s1p1u1p1o1g1u1l1e1c1y1l1e1y1s1u1g1u1a1g1y1u1l 36
21408 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21359
36 ysa1rg1u1Th1r1y1s1g1u1y1s1u1a1s1n1y1s 45
21358 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21330
```

---



17 ysgluAsnGluGluLysAspGluProGluIleGlyIleLysLysGluAa 33  
 |||||  
 51 AAGAAATGAGAGAAAGATGAGCCAGATTGATATAAGAGAGACGT 100  
 34 GlyGluLysArgGluThrLysGluLysGluLys 45  
 |||||  
 101 GGAGAAAAAGAGACAAAGAGAGAAATAG 136

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142752

seq\_documentation\_block:  
 ID AA142752 standard; cDNA; 265 BP.

AC AA142752;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A insert motif.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KM CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52 /tag= a  
 FT /note= "base 52 disrupts the reading frame for  
 FT the translated amino acid sequence given  
 FT in Fig 7"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08144.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Disclosure; Fig 7; 76pp; English.  
 XX  
 CC A composite nucleotide sequence (AA142752) and putative translation  
 CC (AAW08144) sequence are provided of a motif that is found spliced to  
 CC a proportion of chicken CHD-1A clones. The motif is inserted  
 CC between bases 4327 and 4328 of the CHD-1A composite sequence  
 CC (AA142751). None of the 7 CHD-1 clones examined contained the  
 CC complete motif. There are no splice donor or acceptor sites within  
 CC the motif suggesting it is a final rather than an intermediary  
 CC product of splicing. The motif is also found at the 3' end of the  
 CC CHD-W clone CC14 (see also AA142753).  
 XX  
 SQ Sequence 265 BP; 158 A; 16 C; 61 G; 30 T; 0 other;

alignment\_scores:  
 Quality: 180.00 Length: 46  
 Ratio: 4.390 Gaps: 1  
 Percent Similarity: 89.130 Percent Identity: 84.783

alignment\_block:  
 US-08-973-363-14 x AA142752 ..

Align seg 1/1 to: AA142752 from: 1 to: 265

1 AspGlyIleValIserValLysHisProHisLysLysIleLysAlaGlu.L 17  
 |||||  
 1 GATGAGATTGTTTCAGTGAACATCTACATATAAAATATAAGAGAGAA 50  
 17 ysgluAsnGluGluLysAspGluProGluIleGlyIleLysLysGluAa 33  
 |||||  
 51 AAGAAATGAGAGAAAGCCTGAGCCAGATTGATATAAGAGAGACGT 100  
 34 GlyGluLysArgGluThrLysGluLysGluLys 45  
 |||||  
 101 GAAGAAAAAGAGACAAAGAGAGAAATAG 136

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA64139

seq\_documentation\_block:  
 ID AAA64139 standard; DNA; 50000 BP.

AC AAA64139;  
 XX  
 DT 20-DEC-2000 (first entry)  
 XX  
 DE Nucleotide sequence of a beta-tubulin antigen.  
 XX  
 KW Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;  
 KM chronic ear disease; autoimmune disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200050593-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PF 25-FEB-2000; 2000MO-US04795.  
 XX  
 PR 25-FEB-1999; 99US-0121549.  
 XX  
 PA (UYTE-) UNIV TENNESSEE RES CORP.  
 XX  
 PI Yoo TJ;  
 XX  
 DR WPI; 2000-558400/51.  
 XX  
 PT New beta-tubulin antigen in the membranous structure of the inner ear,  
 PT reactive with antibodies of patients with Meniere's disease, for  
 PT diagnosing Meniere's disease and distinguishing this disease from other  
 PT autoimmune ear diseases  
 XX  
 PS Claim 3; Page 51-74; 115pp; English.  
 XX  
 CC The present sequence encodes a beta-tubulin antigen. The protein is  
 CC an antigen of the membranous structure of the inner ear protein, and  
 CC is reactive with antibodies from patients having Meniere's disease.  
 CC Meniere's disease is a chronic ear disease with unknown etiology.  
 CC Serum from patients suffering from this disease contain autoantibodies  
 CC against a 30 kDa cochlear protein antigen. The disease is believed to be  
 CC an autoimmune disease. The beta-tubulin antigen is useful as a target  
 CC substance in diagnosing or detecting Meniere's disease and in  
 CC distinguishing this disease from other autoimmune ear diseases.  
 XX  
 SQ Sequence 50000 BP; 16814 A; 10211 C; 9526 G; 13449 T; 0 other;

alignment\_scores:  
 Quality: 82.00 Length: 39  
 Ratio: 3.154 Gaps: 1  
 Percent Similarity: 66.667 Percent Identity: 51.282

alignment\_block:  
 US-08-973-363-14 x AAA64139 ..

Align seg 1/1 to: AAA64139 from: 1 to: 50000





PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	01-DEC-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251889.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HOMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-483426/52.		
XX			
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
PT	useful for preventing, diagnosing and/or treating cancers and		
PT	metastasis -		
XX			
PS	Disclosure; SEQ ID NO 25592; 3071bp + Sequence Listing; English.		
XX			
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)		
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic		
CC	activity, and can be used in gene therapy and vaccine production. (I)		
CC	proteins and polynucleotides may be used in the prevention, diagnosis and		
CC	treatment of diseases associated with inappropriate (I) expression. For		
CC	example, they may be used to treat disorders associated with decreased		
CC	expression by rectifying mutations or deletions in a patient's genome		
CC	that affect the activity of (I) by expressing inactive proteins or to		
CC	supplement the patients own production of (I). Additionally, (I)		
CC	polynucleotides may be used to produce the secreted (I), by inserting		
CC	the nucleic acids into a host cell and culturing the cell to express the		
CC	protein. (I) proteins and polynucleotides may be used to prevent,		
CC	diagnose and treat immune/hematopoietic-related diseases, especially		
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703		
CC	to AAK87694 represent human immune/hematopoietic antigen genomic		
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		
CC	represent sequences used in the exemplification of the present invention.		
XX			
SQ	Sequence 37783 BP; 9418 A; 9546 C; 9545 G; 9274 T; 0 other;		

Alignment_scores:			
Quality:	81.00	Length:	39
Ratio:	2.793	Gaps:	0
Percent Similarity:	74.359	Percent Identity:	46.154

```
alignment_block:
US-08-973-363-14 x AAK70780/rev .
```

Align seg 1/1 to reverse of: AAK70780 from: 1 to: 37783

7 LysHisProHisLysLysIleLysAlaGluLysGluAsnGluGluLysAs 23  
|||::: |||::: |||::: |||::: |||:::  
16327 AAAGAAAAAACACAGACAGCAGCAAGCAGAAGAGAAAGACAGCAGCA 16278

[illegible]

```
seq_name: /SIDS1/gcgdate/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK76625
seq_documentation_block:
ID   AAK76625 standard; DNA; 37783 BP.
```

AC AAK76625;

DT 07-NOV-2001 (first entry)

DE	Human Immune/haematopoietic antigen genomic sequence	SEQ ID NO: 31437
DE		

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN W0200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0189874.

PR 18-APR-2000; 2000US-0198123.

PR 07-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0337183

PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0338034

PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229287.

PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.

PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513

PR 06-SEP-2000; 2000US-0230437.



|||||  
16227 AGGAGAGGAGAGAG 16211

seq\_name: /SIDSJ/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAK80913

seq\_documentation\_block:

ID AAK80913 standard; DNA; 37763 BP.

XX AAK80913;

DT 07-NOV-2001 (first entry)

XX Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:35725.

KM Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN WO200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228824.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234977.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

17-NOV-2000: 2000US-0249239.  
 PR 17-NOV-2000: 2000US-0249300.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250169.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251858.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 08-DEC-2000: 2000US-0251997.  
 PR 11-DEC-2000: 2000US-0251097.  
 PR 05-JAN-2001: 2001US-0253678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 P1 Rosen CA, Barash SC, Ruben SM;  
 P2  
 P3 WPI, 2001-483426/52.  
 DR  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 P1  
 P2  
 P3  
 P4  
 P5  
 P6  
 P7  
 P8  
 P9  
 P10  
 P11  
 P12  
 P13  
 P14  
 P15  
 P16  
 P17  
 P18  
 P19  
 P20  
 P21  
 P22  
 P23  
 P24  
 P25  
 P26  
 P27  
 P28  
 P29  
 P30  
 P31  
 P32  
 P33  
 P34  
 P35  
 P36  
 P37  
 P38  
 P39  
 P40  
 P41  
 P42  
 P43  
 P44  
 P45  
 P46  
 P47  
 P48  
 P49  
 P50  
 P51  
 P52  
 P53  
 P54  
 P55  
 P56  
 P57  
 P58  
 P59  
 P60  
 P61  
 P62  
 P63  
 P64  
 P65  
 P66  
 P67  
 P68  
 P69  
 P70  
 P71  
 P72  
 P73  
 P74  
 P75  
 P76  
 P77  
 P78  
 P79  
 P80  
 P81  
 P82  
 P83  
 P84  
 P85  
 P86  
 P87  
 P88  
 P89  
 P90  
 P91  
 P92  
 P93  
 P94  
 P95  
 P96  
 P97  
 P98  
 P99  
 P100  
 P101  
 P102  
 P103  
 P104  
 P105  
 P106  
 P107  
 P108  
 P109  
 P110  
 P111  
 P112  
 P113  
 P114  
 P115  
 P116  
 P117  
 P118  
 P119  
 P120  
 P121  
 P122  
 P123  
 P124  
 P125  
 P126  
 P127  
 P128  
 P129  
 P130  
 P131  
 P132  
 P133  
 P134  
 P135  
 P136  
 P137  
 P138  
 P139  
 P140  
 P141  
 P142  
 P143  
 P144  
 P145  
 P146  
 P147  
 P148  
 P149  
 P150  
 P151  
 P152  
 P153  
 P154  
 P155  
 P156  
 P157  
 P158  
 P159  
 P160  
 P161  
 P162  
 P163  
 P164  
 P165  
 P166  
 P167  
 P168  
 P169  
 P170  
 P171  
 P172  
 P173  
 P174  
 P175  
 P176  
 P177  
 P178  
 P179  
 P180  
 P181  
 P182  
 P183  
 P184  
 P185  
 P186  
 P187  
 P188  
 P189  
 P190  
 P191  
 P192  
 P193  
 P194  
 P195  
 P196  
 P197  
 P198  
 P199  
 P200  
 P201  
 P202  
 P203  
 P204  
 P205  
 P206  
 P207  
 P208  
 P209  
 P210  
 P211  
 P212  
 P213  
 P214  
 P215  
 P216  
 P217  
 P218  
 P219  
 P220  
 P221  
 P222  
 P223  
 P224  
 P225  
 P226  
 P227  
 P228  
 P229  
 P230  
 P231  
 P232  
 P233  
 P234  
 P235  
 P236  
 P237  
 P238  
 P239  
 P240  
 P241  
 P242  
 P243  
 P244  
 P245  
 P246  
 P247  
 P248  
 P249  
 P250  
 P251  
 P252  
 P253  
 P254  
 P255  
 P256  
 P257  
 P258  
 P259  
 P260  
 P261  
 P262  
 P263  
 P264  
 P265  
 P266  
 P267  
 P268  
 P269  
 P270  
 P271  
 P272  
 P273  
 P274  
 P275  
 P276  
 P277  
 P278  
 P279  
 P280  
 P281  
 P282  
 P283  
 P284  
 P285  
 P286  
 P287  
 P288  
 P289  
 P290  
 P291  
 P292  
 P293  
 P294  
 P295  
 P296  
 P297  
 P298  
 P299  
 P300  
 P301  
 P302  
 P303  
 P304  
 P305  
 P306  
 P307  
 P308  
 P309  
 P310  
 P311  
 P312  
 P313  
 P314  
 P315  
 P316  
 P317  
 P318  
 P319  
 P320  
 P321  
 P322  
 P323  
 P324  
 P325  
 P326  
 P327  
 P328  
 P329  
 P330  
 P331  
 P332  
 P333  
 P334  
 P335  
 P336  
 P337  
 P338  
 P339  
 P340  
 P341  
 P342  
 P343  
 P344  
 P345  
 P346  
 P347  
 P348  
 P349  
 P350  
 P351  
 P352  
 P353  
 P354  
 P355  
 P356  
 P357  
 P358  
 P359  
 P360  
 P361  
 P362  
 P363  
 P364  
 P365  
 P366  
 P367  
 P368  
 P369  
 P370  
 P371  
 P372  
 P373  
 P374  
 P375  
 P376  
 P377  
 P378  
 P379  
 P380  
 P381  
 P382  
 P383  
 P384  
 P385  
 P386  
 P387  
 P388  
 P389  
 P390  
 P391  
 P392  
 P393  
 P394  
 P395  
 P396  
 P397  
 P398  
 P399  
 P400  
 P401  
 P402  
 P403  
 P404  
 P405  
 P406  
 P407  
 P408  
 P409  
 P410  
 P411  
 P412  
 P413  
 P414  
 P415  
 P416  
 P417  
 P418  
 P419  
 P420  
 P421  
 P422  
 P423  
 P424  
 P425  
 P426  
 P427  
 P428  
 P429  
 P430  
 P431  
 P432  
 P433  
 P434  
 P435  
 P436  
 P437  
 P438  
 P439  
 P440  
 P441  
 P442  
 P443  
 P444  
 P445  
 P446  
 P447  
 P448  
 P449  
 P450  
 P451  
 P452  
 P453  
 P454  
 P455

```

seq_documentation_block:
ID      ABA46373 standard: DNA; 420 BP.
XX
XX
XX      ABA46373;
XX
XX      01-FEB-2002 (first entry)
XX
XX      Human breast cell single exon nucleic acid probe #5068.
XX
XX      Human; microarray; single exon probe; gene expression; breast;
XX      disease; cancer; ss.
XX
XX      Homo sapiens.
XX
XX      MO200157271-A2.
XX
XX      09-AUG-2001.
XX
XX      30-JAN-2001; 2001WO-US00662.
XX
XX      04-FEB-2000; 2000US-0180312.
XX      26-MAY-2000; 2000US-0207456.
XX      30-JUN-2000; 2000US-0608408.
XX      03-AUG-2000; 2000US-0632366.
XX      21-SEP-2000; 2000US-0234687.
XX      27-SEP-2000; 2000US-0236359.
XX      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX      WPI; 2001-496933/54.
XX
XX      New spatially-addressable set of single exon nucleic acid probes,
XX      useful for measuring gene expression in sample derived from human
XX      breast, comprises number of single exon nucleic acid probes -
XX
XX      Claim 1: SEQ ID NO 5068; 327bp + sequence listing; English.
XX
XX      The invention relates to a spatially-addressable set of single exon
XX      nucleic acid probes for measuring gene expression in a sample derived
XX      from human breast and BT 474 cells. The method involves contacting
XX      the probes with a collection of detectably labelled nucleic acids
XX      derived from mRNA of human breast, and then measuring the label
XX      bound to each probe of the microarray. The probes are useful for
XX      verifying the expression of regions of genomic DNA predicted to
XX      encode proteins. They are useful for gene discovery, and for
XX      determining predisposition and/or prognosing breast disease. Gene
XX      expression analysis is useful for assessing the toxicity of chemical
XX      agents on cells. The microarray of this invention presents a far greater
XX      diversity of probes for measuring gene expression, with far less bias
XX      than expressed sequence tag microarrays. The method is suitable for
XX      rapid production of functional information from genomic sequence. The
XX      present sequence is a single exon nucleic acid probe of the invention.
XX      Note: The sequence data for this patent did not form part of the
XX      printed specification, but was obtained in electronic format directly
XX      from WIPO at fp.wipo.int/pub/published\_pct\_sequences.
XX
XX      Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment_scores:
Quality: 79.00 Length: 45
Ratio: 2.469 Gaps: 2
Percent Similarity: 71.111 Percent Identity: 42.222

alignment_block:
US-08-973-363-14 x ABA46373 ..

Align seg 1/1 to: ABA46373 from: 1 to: 420
4 ValSerValIySHISPRHIS.....LylIyIyISAlIySAlIyIy 17

```

```

228 ATAACTTACGTGATCCTCATTTCAAGAGAAAGAGAGATTGAGACA 277
      ::::::::::: |||||  ' |||||  |||  ::::|
17  sgluasnglululsaspgluProgluileglylleylslsglualag 34
      |||:::|||||:::  ::|:::
278 GGAGAGAAAGAGAGAGATGAGAGAAA.....AACAAAGAAAGACA 318
      |||:::||||  ::|::|::|::|::|::|::|
34  lylulysarggluThrlsyclulysgluasnllys 45
      |||:::||||  ::|::|::|::|::|::|::|
319 AGGAGAAAGAGAGAGAAAATGAGAAAGGAGAAAG 353

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA56939

seq\_documentation\_block:

ID ABA56939 standard; DNA; 420 BP.

XX ABA56939;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #5244.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX MO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver.

XX Claim 1; SEQ ID NO 5244; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment\_scores:

Quality: 79.00 Length: 45

Ratio: 2.469 Gaps: 2

Percent Similarity: 71.111 Percent Identity: 42.222

alignment\_block:

US-08-973-363-14 x ABA56939 ..

Align seg 1/1 to: ABA56939 from: 1 to: 420

```

4  ValserVallySHisProHis.....LysLysLleySAlaGluLyl 17
      ::::::::::: |||||  |||||  |||||  ::::|
228 ATAACTTACGTGATCCTCATTTCAAGAGAAAGAGAGATTGAGACA 277
      |||:::|||||:::  ::|:::
17  sgluasnglululsaspgluProgluileglylleylslsglualag 34
      |||:::|||||:::  ::|:::
278 GGAGAGAAAGAGAGAGATGAGAGAAA.....AACAAAGAAAGACA 318
      |||:::||||  ::|::|::|::|::|::|::|
34  lylulysarggluThrlsyclulysgluasnllys 45
      |||:::||||  ::|::|::|::|::|::|::|
319 AGGAGAAAGAGAGAGAAAATGAGAAAGGAGAAAG 353

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA26551

seq\_documentation\_block:

ID ABA26551 standard; DNA; 420 BP.

XX ABA26551;

XX 23-JAN-2002 (first entry)

XX Probe #5017 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX MO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts.

XX Claim 1; SEQ ID NO 5017; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment\_scores:

Quality: 79.00 Length: 45

[illegible]

```

XX alignment_scores:                                     Quality:   79.00      Length:   45
XX                                         Ratio:     2.469
XX Percent Similarity:    71.111      Percent Identity:  42.222
XX
XX alignment_block:
XX US-08-973-363-14 x AAK05034      ..
XX
XX Align seg 1/1 to: AAK05034 from: 1 to: 420
XX
XX         4 valserValLysHisProHis.....LysLysIleLysAlaGluTyr 17
XX          ::::::::::::::::::::| | | | | | | | | | | | | | | |
XX 228 ATTAACCTPACGTGATCCTCATTTCAAGGAGAAGAAAGAAATTGAAGAA 277
XX
XX 17 sgluaSngLuGlulYulYsaSpGuLuproGluIleGlyIleLysLysGluInlaG 34
XX  |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
XX 278 GGGAGGAGAAAGAGAGAAAGATGAGAACAAA.....AACAAAGAGAGGA 318
XX
XX 34 LysIulYsaArgIuThrLysGluLysGluYsaInlYs 45
XX  |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
XX 319 AGGAGAGAGAGAGAGAAATAATGAGAGGAGGAAGAG 353
XX
XX seq_name: /SIDB1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK30564
XX
XX seq_documentation_block:
XX ID AAK30564 standard; DNA: 420 BP.
XX
XX AC AAK30564;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 5121.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001MO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX XX
XX XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
XX
XX PT Humm genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX XX Example 4; SEQ ID NO: 5121; 65bp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX QQ Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

```















```
17 sgluAsngluGluLysAspGluProGluIleGlyIleLysLysGluAlaG 34
      ::::::::::::::::::::
82 AGACAGTGAAGAAAAACCTGACGACATGTTATATAAGAAAGAACAG 33
      ::::::::::::::::::::
34 GlyLysArgGluThrLysGluLysGluAla 44
      ::::::::::::::::::::
32 AAGAAAGAGGGAAGCAAAAGAAAGAGAGAT 1
      ::::::::::::::::::::

seq_name: gb_est2:BF239967

seq_documentation block:
LOCUS       BF239967               821 bp      mRNA           linear      EST 14-NOV-2000
DEFINITION  601905170F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4133129 5',
            mRNA sequence.
ACCESSION   BF239967
VERSION     BF239967.1  GI:11153890
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 821)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: CLONTECH Laboratories, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            http://image.llnl.gov
            Plate: L10M1033 row: k column: 18
            High quality sequence stop: 562.
            Location/Qualifiers
                source
                    1. 821
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_image="4133129"
                        /clone_lib="NIH_MGC_54"
                        /tissue_type="from chronic myelogenous leukemia"
                        /lab_host="DH10B (T1 phage-resistant)"
                        /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                        Site:1: Sfil (ggcgccgcgcgcgc); Site:2: Sfil (ggcgccatgagc
                        ); Double-stranded cDNA was prepared from cell line RNA.
                        5' and 3' adaptors were used in cloning as follows: 5'
                        adaptor sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor
                        sequence: 5'-ATTCTAGAGCGCGACGCGACATG-dT(30)BN-3'
                        (where B = A, C, G or G and N = A, C, G, or T). Average
                        insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                        contained inserts by PCR. This library was enriched for
                        full-length clones and was constructed by Clontech
                        Laboratories (Palo Alto, CA)."
```

```
17 sglu.AsngluGluLysAspGluProGluIleGlyIleLysLysGluAla 33
      ::::::::::::::::::::
284 AGACAGTGAAGAAAAACCTGACGACATGTTATATAAGAAAGAACCA 333
      ::::::::::::::::::::
34 GlyLysArgGluThrLysGluLysGluAla 45
      ::::::::::::::::::::
334 GAAGAAAAAGGGAAGCAAAAGAAAGAGATATAA 369
      ::::::::::::::::::::

seq_name: gb_est1:AW387264

seq_documentation block:
LOCUS       AW387264               249 bp      mRNA           linear      EST 04-FEB-2000
DEFINITION  MRL-ST0088-101199-003-b09 ST0088 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW387264
VERSION     AW387264.1  GI:6891923
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 249)
AUTHORS    HCGP http://www.ludwig.org.br/ORESTES
TITLE      The FAPESP/LICR Human Cancer Genome Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?l=MRL&t=MRL-ST0088-
            101199-003-b09&t3=1999-11-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 27
            High quality sequence stop: 248.
            Location/Qualifiers
                source
                    1. 249
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="ST0088"
                        /dev_stage="Adult"
                        /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                        Site_2: SmaI; A mini-library was made by cloning products
                        derived from ORESTES PCR (U.S. Letters Patent application
                        No. 196,716 - Ludwig Institute for Cancer Research)
                        profiles into the puc 18 vector. Reverse transcription of
                        tissue mRNA and cDNA amplification were performed under
                        low stringency conditions."
```

109 AGACAGTGAAGAAAAACCTGACCATGTTTATATATAAGAGAACACAG 60  
 34 LysGluLysArgGluThrLysGluLysGluAsnLys 45  
 |||||.....  
 59 MAGCAAGAGGAGGACACAGCAAGAACGAACTCA 25  
 seq\_name: gb\_gss:BH055497

seq\_documentation\_block: 653 bp DNA linear GSS 18-JUL-2001  
 LOCUS BH055497  
 DEFINITION RPI-24-278116.TJ RPI-24 Mus musculus genomic clone RPI-24-278116  
 , DNA sequence.  
 ACCESSION BH055497  
 VERSION BH055497.1 GI:14862403  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 653)  
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akireti, B., Levins, M.,  
 Tsengaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregorgis, E.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPI-24  
 Unpublished (1999)  
 Other GSSs: RPI-24-278116.TJ  
 Contact: Shaying Zhao  
 Department of Biomedical Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@tigr.org). Clones may be purchased from BACPAC  
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)  
 Plate: 278 row: 1 column: 16  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..653  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_id="RPI-24-278116"  
 /clone\_id="RPI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 255 a 101 c 172 g 125 t  
 ORIGIN

alignment\_scores: Quality: 90.00 Length: 39  
 Ratio: 3.333 Gaps: 0  
 Percent Similarity: 69.231 Percent Identity: 51.282

Alignment block:  
 US-08-973-363-14 x BH055497 ..  
 Align seg 1/1 to: BH055497 from: 1 to: 653

7 LysHisProHisLysLysIleLysAlaGluLysGluAsnGluGluLysAs 23  
 |||||.....  
 140 AAG 189

23 pgiuprogliueglyllyllysllysluaglyllysluysarggluthtl 40  
 |||.....  
 190 AGAAGAAAGATAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTGA 239  
 40 ysgLysGluAsnLys 45  
 |||||.....  
 240 AAGAAAAAGAAAAACAAA 256  
 seq\_name: gb\_est2:BF068827

seq\_documentation\_block: 488 bp mRNA linear EST 06-DEC-2001  
 LOCUS BF068827  
 DEFINITION s103c11.y1 Gm-cl065 glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl065-237 5' similar to TR:003982 003982 HYPOTHEICAL 19.8 KD  
 PROTEIN. mRNA sequence.  
 ACCESSION BF068827  
 VERSION BF068827.1 GI:10845778  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine  
 1 (bases 1 to 488)  
 Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Corvelli, V., Khana  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Stepcie, M., Theising, B., Allen, M., Bowers  
 J., Person, B., Swaller, T., Giddons, M., Pape, D., Harvey, N., Schurk  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 Public soybean EST Project  
 Unpublished (1999)  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Resgen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntville, AL 35801 For further information  
 call: (800)-533-4363 or contact via email: [custresgen.com](mailto:custresgen.com)  
 Insert Length: 822 Std Error: 0.00  
 High quality sequence stop: 414.

FEATURES  
 source Location/Qualifiers  
 1..488  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl065-237"  
 /clone\_id="Gm-cl065"  
 /tissue\_type="germinating shoots"  
 /lab\_host="DH10B"  
 /note="Vector: Bluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI. The cDNA library was constructed from mRNA isolated  
 germinating shoots of the cultivar Williams. The seeds  
 were allowed to germinate for 24 hours prior to being  
 cold stressed for 2 days at 4C. Complementary DNA was  
 synthesized from mRNA using a primer consisting of a  
 poly(dT) sequence with a XhoI restriction site. EcoRI  
 adapters were ligated to the blunt-ended cDNA fragments  
 followed by XhoI digestion. The cDNA fragments were  
 directionally cloned into the EcoRI-XhoI restriction site  
 of the pBluescript vector. The ligated cDNA fragments were  
 transformed into DH10B host cells (GibcoBRL). This library  
 was constructed in the laboratory of Dr. Randy  
 Shoemaker."

BASE COUNT 180 a 73 c 115 g 120 t  
 ORIGIN

alignment\_scores: Quality: 85.50 Length: 45

Ratio: 2.672 Gaps: 2  
Percent Similarity: 71.111 Percent Identity: 46.667

Alignment\_block:  
US-08-973-363-14 x BF068827 ..

Align seg 1/1 to: BF068827 from: 1 to: 488

```

6 VallyshisProHis.....LysylsiletyalaGluLulY 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
38 GTGAAGGAGCCTGAGAGAAAGAGAGATCCCAAGAAAGTAGAGCAAAAGA 87
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 sGLuAnGluGluLulYsAspGluProGluIleGlyIleLysylsGluAaG 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
88 AGAGGAGGAGAAAGAAAGAGAGAGGGA.....AAGAAGCAAGAAAG 125
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
34 lYGLuLYsARGlUthrlYsGluLYsGluAnLYs 45
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
126 GGAAGAAAGAAAGAGAGAGAGATGAGAGAAAG 160

```

seq\_name: gb\_est2:BG790466

seq\_documentation\_block:

LOCUS BG790466 537 bp mRNA linear EST 29-NOV-2001  
DEFINITION sae69h05.y1 Gm-cl064 Glycine max cDNA clone GENOME SYSTEMS CLONE  
ID: Gm-cl064-3801 5' similar to TR:003982 003982 HYPOTHETICAL 19.8  
KD PROTEIN.; mRNA sequence.

ACCESSION BG790466  
VERSION BG790466.1 GI:14126028

KEYWORDS EST.  
SOURCE soybean.

ORGANISM Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.  
REFERENCE 1 (bases 1 to 537)  
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Fleising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., Mccann,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: Resgen, Invitrogen Corp. 2130  
South Memorial Parkway Huntville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
High quality sequence stop: 421.

#### FEATURES

source

1. 537  
Location/Qualifiers  
/organism="Glycine max"  
/cultivar="Williams"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl064-3801"  
/clone\_1lb="Gm-cl064"  
/tissue\_type="seedling epicotyls"  
/dev\_stage="2 week old"  
/lab\_host="DH10B"  
/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2: XhoI; The cDNA library was constructed from mRNA isolated from the epicotyls of 2 week old seedling for the cultivar Williams. The seedlings were germinated in a growth chamber, excised above the soil level, and the plants were placed in a 100 ppm solution of auxin for 24 hours prior to harvesting. Complementary DNA was synthesized

from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 193 a 93 c 126 g 125 t  
ORIGIN

alignment\_scores:  
Quality: 85.50 Length: 45  
Ratio: 2.672 Gaps: 2  
Percent Similarity: 71.111 Percent Identity: 46.667

alignment\_block:  
US-08-973-363-14 x BG790466 ..

Align seg 1/1 to: BG790466 from: 1 to: 537

```

6 VallyshisProHis.....LysylsiletyalaGluLulY 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
343 GTGAAGGAGCCTGAGAGAAAGAGAGATCCCAAGAAAGTAGAGCAAAAGA 392
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
17 sGLuAnGluGluLulYsAspGluProGluIleGlyIleLysylsGluAaG 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
393 AGAGGAGGAGAAAGAAAGAGAGGGA.....AAGAAGCAAGAAAG 430
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
34 lYGLuLYsARGlUthrlYsGluLYsGluAnLYs 45
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
431 GGAAGAAAGAAAGAGAGAGAGATGAGAGAAAG 465

```

seq\_name: gb\_est2:BE822468

seq\_documentation\_block:

LOCUS BE822468 796 bp mRNA linear EST 24-MAY-2001  
DEFINITION GW700017B10E2 Gm-r11070 Glycine max cDNA clone Gm-r11070-6747 3',  
mRNA sequence.

ACCESSION BE822468  
VERSION BE822468.1 GI:10254702  
KEYWORDS EST.  
SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.  
REFERENCE 1 (bases 1 to 796)  
AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelting,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.  
A Functional Genomics Program for Soybean (NSF 9872565)  
Unpublished (1999)  
Other\_ESTS: AW277969 corresponding to Gm-cl019-3195 (5')  
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)  
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics

University of Illinois  
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA  
Tel: (217) 244-6147  
Fax: (217) 333-4582  
Email: l-vodkin@uiuc.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134. For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or info@genome  
systems.com web site: www.genomesystems.com  
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

#### FEATURES

source

1. 796  
Location/Qualifiers



```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-6747"
/clone_1bp="Gm-r1070"
/Note="The library Gm-r1070 is a sequence-driven, rerecked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones, or
the different libraries, was used to select singletons, or
a representative of each contig, which were rerecked to
form library Gm-r1070. The cDNA clones of the rerecked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the Laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
. Rerecking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Kock Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/kock.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."
BASE COUNT      208 a      177 c      107 g      274 t      30 others
ORIGIN

Alignment_scores:
  Quality:      84.00      Length:      39
  Ratio:        2.897      Gaps:      1
Percent Similarity: 74.359      Percent Identity: 48.718

alignment_block:
US-08-973-363-14 x BE822468/rev ..

Align seg 1/1 to reverse of: BE822468 from: 1 to: 796

7 LysHisProHisIleLysIleLysAlaGluLysGluAsnGluGluLysAs 23
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 AAGAGAGATCCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
23 pGIUPrGtIleGlyIleLysLysGluAlaGlyLysArgGluThrL 40
||||| ||||| ||||| ||||| ||||| ||||| |||||
577 GGAGGGA.....ACGAGAGAGAGAGAGAGAGAGAGAGAGAGA 540
40 ysgIuLysGluAsnLys 45
||||| ||||| ||||| ||||| ||||| ||||| |||||
539 AAGAGAGATGAGAGAGAG 523

seq_name: gb_estl:BE211196

seq_documentation_block:
LOCUS      BE211196      294 bp      mRNA      linear      EST 04-DEC-2001
DEFINITION      sos8h05.y1 Gm-cl039 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl039-2194 5' similar to TR:003982 003982 HYPOTHETICAL 19.8 KD
PROTEIN. // mRNA sequence.
ACCESSION      BE211196
VERSION        BE211196
KEYWORDS       BE211196.1 GI:8827466
SOURCE         soybean
ORGANISM       Glycine max
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
               Glycine
REFERENCE      1 (bases 1 to 294)
AUTHORS       Shoemaker,R., Keim,P., Vodkin,L., Erpelnding,J., Coryell,V., Khanna

```

```

A. Bolla,B., Marra,M., Hallier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,R., Steptoe,M., Theising,B., Allen,M., Bowers
,I., Pearson,B., Swales,T., Gibbons,M., Pepe,D., Harvey,N., Schurk
,R., Ritten,E., Kohn,S., Shin,T., Jackson,I., Cardenas,W., McCann
,R., Mattern,R. and Wilson,R.
Public soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: curesgen.com
Insert length: 1211 Std Error: 0.00
High quality sequence stop: 245.
Location/Qualifiers
1..294
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl039-2194"
/clone_1bp="Gm-cl039"
/tissue_type="whole seedling without cotyledons"
/lab_host="DH10B"
/Note="Vector: BluescriptII SK+; Site 1: EcoRI; Site 2:
XhoI; This cDNA library was constructed from mRNA isolated
from 2 week old seedlings with the cotyledons removed at
the time of harvest. The seedlings for the cultivar Oden
were grown in a growth chamber using germination paper.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with a XhoI restriction
site. EcoRI adapters were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA fragments
were directionally cloned into the EcoRI-XhoI restriction
site of the Bluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (Gibco BRL). This
library was constructed by Dr. Randy Shoemaker."
BASE COUNT      130 a      49 c      78 g      37 t
ORIGIN

Alignment_scores:
  Quality:      83.00      Length:      39
  Ratio:        2.677      Gaps:      1
Percent Similarity: 79.487      Percent Identity: 46.154

alignment_block:
US-08-973-363-14 x BE211196 ..

Align seg 1/1 to: BE211196 from: 1 to: 294

7 LysHisProHisIleLysIleLysAlaGluLysGluAsnGluGluLysAs 23
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40 GAGGAGCCCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 89
23 pGIUPrGtIleGlyIleLysLysGluAlaGlyLysArgGluThrL 40
||||| ||||| ||||| ||||| ||||| ||||| |||||
90 AGAGCCT.....ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 127
40 ysgIuLysGluAsnLys 45
||||| ||||| ||||| ||||| ||||| ||||| |||||
128 AAGAGAGAGAGAGAGAG 144

seq_name: gb_estl:AM459832

seq_documentation_block:
LOCUS      AM459832      470 bp      mRNA      linear      EST 03-DEC-2001
DEFINITION      shs908.y1 Gm-cl016 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl016-7959 5' similar to TR:003982 003982 HYPOTHETICAL 19.8 KD
PROTEIN. // mRNA sequence.

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:41:11 ; Search time 10310.5 Seconds  
(without alignments) 310.532 Million cell updates/sec

Title: US-08-973-363-2

Sequence: 1 ATTCTTCAGATGATGCTGA.....CTCAGAGACTTGTGTGCG 153

Scoring table: OLIGO\_MNC  
Gapop 60.0, Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

GenEmb1:\*  
1: gb\_da:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_com:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pi:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_jm:\*  
20: em\_lm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	153	100.0	153	6	A58683
2	102	66.7	5349	10	MUSC83
3	54	35.3	1311	6	AR029026
4	27	17.6	5417	9	AR029026
5	27	17.6	101220	9	AC092272
6	27	17.6	134365	9	AC012624
7	27	17.6	145079	2	AC021449
8	27	17.6	145639	2	AC008531
9	27	17.6	193446	2	AC091946
10	27	17.6	219258	2	AC022121
11	27	17.6	276181	2	AC092382
12	25	16.3	153	6	A58684
13	25	16.3	2292	5	D14316
14	25	16.3	6608	6	A58691
15	25	16.3	6872	5	AF004397
16	21	13.7	136229	2	AC106847
17	20	13.1	153	6	A58685
18	19	12.4	16797	4	MIHGGC
19	19	12.4	81035	8	AB025631
20	19	12.4	86606	2	AC097591
21	19	12.4	98507	2	AC096218
22	19	12.4	110000	2	LMFLCHR18_01
23	19	12.4	112664	2	AC008400
24	19	12.4	154859	9	AC012596
25	19	12.4	156280	9	AC079595
26	19	12.4	156563	2	AC079458
27	19	12.4	160690	2	AC023457
28	19	12.4	160894	2	AL450462
29	19	12.4	167466	2	AC084117
30	19	12.4	175357	2	AC079775
31	19	12.4	182770	2	AL596210
32	19	12.4	190000	2	AC004580
33	19	12.4	198757	2	AC073661
34	19	12.4	202540	2	AC009483
35	19	12.4	206589	9	AC009483
36	19	12.4	210256	2	AC108116
37	19	12.4	210515	9	AC097382
38	19	12.4	210636	9	AC006443
39	19	12.4	215705	9	AL589931
40	19	12.4	216438	2	AL359881
41	18	11.8	348	5	AF060702
42	18	11.8	380	3	AF173269
43	18	11.8	418	3	AF062983
44	18	11.8	2266	9	BC017739
45	18	11.8	10591	1	AB006096

## ALIGNMENTS

RESULT	1	153 bp	DNA	linear	PAT 06-MAR-1998
A58683	A58683	Sequence 2 from Patent WO9639505.			
LOCUS	A58683	153 bp	DNA	linear	PAT 06-MAR-1998
DEFINITION	A58683	Sequence 2 from Patent WO9639505.			
ACCESSION	A58683	GI:3714246			
VERSION	A58683.1	GI:3714246			
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 153)				
AUTHORS	Giuffridis, R. and Tiwari, B.				
TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS				
JOURNAL	Patent: WO 9639505-A 2 12-DEC-1996;				
COMMENT	ISIS INNOVATION (GB)				
FEATURES	Other publication AU 5906996 961224.				
source	Location/Qualifiers				
	1..153				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT	61 a	37 c	27 g	28 t	
ORIGIN					



JOURNAL 97470991 Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
REFERENCE 2 (bases 1 to 5947)  
AUTHORS Moodabe T  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
FEATURES  
source  
1. 5947  
/organism="Homo sapiens"  
/db.xref="taxon:9606"  
/chromosome="5"  
/map="9q15-21; near WI-5811"  
1. 5947  
/gene="CHD1"  
164. 5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db.xref="GI:2645429"  
/translation="MNGHSDSEYRNSGSESSQSDSDSDSGASGSGSGSSGSSSDGSS  
SOGSSSDSDSGSGSGSESDTSRENNVQAKPKYDGAEFKSSPSILAVQKRALI  
KQOQOQOQOQOQOASNSGSESDSSSEVKKRKHDEDMQSGSGSPSS  
GSDSESEERKSSCDETSDEYERKNNKSRKPKMRKSGKGLIGOKKQIDISE  
DDDEDYDNDKRSRRQATVAVYKDEEMKTSDDILEYCGDYPOPEEEFEETIR  
FMDCRIGKAGATTTIYAVEADGDPNAGFKENKEGEIOLYIKMGSHIHNTMT  
EETLKQONVGMKKLDNYKKKDOETKRMKNASPEDYEVNCOQELTDLKHQOIVG  
RIASHNOKAAGPYDYCKMKGGLPYSECSDGALISKKFOACIDEFSPNSKSTTP  
FRCKVVKORPREVALKKPKYIGGHEGLDLDQNLGKLMHSMGSGNSCLADIM  
GLGKTITTSFLNLYFHHOLYGPFLVPLSTLSMRQIOWASQONNAVYLGIDN  
SNMRTTTEHTHHOKRLKEFVILITTYELLKDKAFIGGNMAFIGVDEARLKNDSS  
LLYTLIDESPNNRLLTGPCLONSLEKMLSHIHPKESWMDPEEPHKGKRG  
YASLHKEFLPRLPRVKDVEKSLPAKVQITLPMKSNALOKOYKXKTLIRYKRLSG  
SKGSTSEFLIMWELKQCNCHYIKPPNMFYKQDAOH1BSKSLITLIDKLT  
PLPRGRGRLITSOMVRLDILATYIKYKQPROLOGSGTCGTRKQALDPNNKESB  
DCELTSTPAGGGLGINTASDVTYIDSSNNQNLQANQARTRKQKONITRYL  
KGSVEDILFRAKKNVLDHLYQKNDTGKLYHLSASSTSPKNGEISALIKG  
ABELFKPDEDEQEDIDELIKRAEIHENREPELTVDELISQFVNAFNMWDD  
DELEFPRNNKNEBELTEPDORRLDEERQKLELELTVDELPRNNKQKISFNSBER  
KSKRTISGSDSDSISEGRKPKKGRRTIPRENKGSDELTRRTIKSTKRGPLE  
RLDIAADALVYSKSELDRLGLGVHNCIKALDSSSTETGRLGKVGKTFRI  
SGOVNKLIVTSHHEELPLPKRTSPDPERKQTTIPCHTKAHDPIDMGKEDSNL  
IGIYEGYGSWEMIKMDPDLSTHKLILPDPPKPKQAKQLOTADYLILKLSRLAK  
EALSGAGSSRRKARAKNNKAMSIVYKEELISDSSPLSEKDEDDKLSSESKSDGR  
ERSKSSVDAVPHITAGSEFPISESESELDQKTFICKERNRPVKALKQDPRPK  
GLSERQLEHTROCLIKITGHTTECHKEVTNPEDQIKERNMLIYVSKFTFDRKLI  
KLYHAIKRRQESQNSDQNSNLPVYINPDEVLKENTHDDSSRDSYSDHRLIQ  
YHDHAKRHOGDSYKRSKSRKRPYSFSGKDHDMVHYQDSRYYSDBREKRLDHD  
RSRHRSNLEGSLSKDRSHSHDRHLSHRSSESYTHNKSMDYRHSDMOMDH  
RASSGGRSLDQRYGSRSPFEHVSHEKSTPEHTWSRKT"

BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 5947;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCGGTGACAGACTACCTATCAAA 102  
|||||  
Db 4070 CAGACCGGTGACAGACTACCTATCAAA 4096

RESULT 5  
AC092372 101220 bp DNA linear PRI 07-DEC-2001  
LOCUS Homo sapiens chromosome 5 clone Rpl1-58M12, complete sequence.  
DEFINITION AC092372  
ACCESSION AC092372.3 GI:17402768  
VERSION  
KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Dec 7, 2001 this sequence version replaced gi:15290448.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146.7kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.  
Location/Qualifiers  
1. 101220  
/organism="Homo sapiens"  
/db.xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-58M12"

BASE COUNT 34122 a 18862 c 17827 g 30409 t  
ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 101220;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCGGTGACAGACTACCTATCAAA 102  
|||||  
Db 26887 CAGACCGGTGACAGACTACCTATCAAA 26861

RESULT 6  
AC012624 134365 bp DNA linear PRI 21-JUL-2001  
LOCUS Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.  
ACCESSION AC012624  
VERSION AC012624.6 GI:14993679  
KEYWORDS  
HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission  
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.  
FEATURES  
    source  
        1..134365  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="5"  
            /clone="CTD-2082117"  
BASE COUNT 40414 a 24497 c 25503 g 43951 t  
ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 134365;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
          |||||  
Db 118292 CAGACCCGTCAGACTACCTCATCAAA 118318

RESULT 7  
AC021449 143079 bp DNA linear HTG 10-SEP-2000  
LOCUS Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
DEFINITION pieces.  
AC021449  
VERSION AC021449.3 GI:10047806  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-58M12  
Unpublished  
2 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerity,R., Beda,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,  
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Deaellano,K., Dewar,K., Domino,M., Doyle,M., Fenesfor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Menais,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,Y., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Sudramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:1407963.  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5154  
Center clone name: 58\_M\_12  
----- Summary Statistics

Sequencing vector: M13; M77815, 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 14400; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.  
FEATURES  
    source  
        1..143079  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /clone="RP11-58M12"  
            /clone\_lib="RPCI-11 Human Male BAC"  
            1..38820  
                /note="assembly\_fragment"  
                clone\_end:386  
                vector\_side:left  
            38921..40411  
                /note="assembly\_fragment"  
                40512..43279  
                /note="assembly\_fragment"  
                43380..46905  
                /note="assembly\_fragment"  
                47006..51830  
                /note="assembly\_fragment"  
                51931..62619  
                /note="assembly\_fragment"  
                62720..75408  
                /note="assembly\_fragment"  
                75509..92516  
                /note="assembly\_fragment"  
                92617..106409  
                /note="assembly\_fragment"  
                106510..143079  
                /note="assembly\_fragment"  
                clone\_end:17  
                vector\_side:right  
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 17.6%; Score 27; DB 2; Length 143079;



Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 116192 CAGACCCGTCAGACTACCTCATCAAA 116218

## RESULT 8

AC008531 145659 bp DNA linear HTG 14-FEB-2001  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.  
AC008531 3 GI:12830078  
VERSION  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE  
ORGANISM human.

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 145659)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
AUTHORS Unpublished  
2 (bases 1 to 145659)  
TITLE  
REFERENCE DOE Joint Genome Institute.  
JOURNAL Direct Submission

COMMENT  
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced g1:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11  
-----

Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 56174: contig of 56174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 113327: gap of unknown length  
\* 113328 113427: contig of 4963 bp in length  
\* 113428 113530: gap of unknown length  
\* 113531 113634: contig of 1404 bp in length  
\* 113635 113734: gap of unknown length  
\* 113735 123297: contig of 3503 bp in length  
\* 123298 123398: gap of unknown length  
\* 123399 145659: contig of 22262 bp in length.  
Location/Qualifiers  
1. 145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="5"  
/clone="CTC-480B11"  
/clone\_lib="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 17.6%; Score 27; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 46815 CAGACCCGTCAGACTACCTCATCAAA 46841

## RESULT 9

AC091946 193446 bp DNA linear HTG 09-JUN-2001  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.  
AC091946  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE  
ORGANISM human.

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 193446)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
AUTHORS Unpublished  
2 (bases 1 to 193446)  
TITLE  
REFERENCE DOE Joint Genome Institute.  
JOURNAL Direct Submission

COMMENT  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RP11-36012  
-----

Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-gel estimation  
Estimated insert size: 190246; sum-of-contigs estimation  
Quality coverage: 5.79 in Q20 bases; agarose-gel estimation  
Quality coverage: 5.38 in Q20 bases; sum-of-contigs estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2668: gap of unknown length  
\* 2669 3823: contig of 1056 bp in length  
\* 3824 3823: gap of unknown length  
\* 3824 5920: contig of 1997 bp in length  
\* 5921 6021: gap of unknown length  
\* 6022 7494: contig of 1474 bp in length  
\* 7495 7595: gap of unknown length  
\* 7596 9096: contig of 1502 bp in length

\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12460: contig of 2100 bp in length  
\* 12461 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14711 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18069: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20534: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 23615: gap of unknown length  
\* 23616 27563: contig of 3948 bp in length  
\* 27564 27663: gap of unknown length  
\* 27664 30987: contig of 3324 bp in length  
\* 30988 31087: gap of unknown length  
\* 31088 36065: contig of 4978 bp in length  
\* 36066 36165: gap of unknown length  
\* 36166 40978: contig of 4813 bp in length  
\* 40979 41078: gap of unknown length  
\* 41079 45663: contig of 4385 bp in length  
\* 45664 45763: gap of unknown length  
\* 45764 51745: contig of 5982 bp in length  
\* 51746 51845: gap of unknown length  
\* 51846 57359: contig of 5514 bp in length  
\* 57360 57459: gap of unknown length  
\* 57460 67881: contig of 10422 bp in length  
\* 67882 67981: gap of unknown length  
\* 67982 74132: contig of 6151 bp in length  
\* 74133 74232: gap of unknown length  
\* 74233 79795: contig of 5563 bp in length  
\* 79796 79895: gap of unknown length  
\* 79896 87511: contig of 7616 bp in length  
\* 87512 87611: gap of unknown length  
\* 87612 92791: contig of 5180 bp in length  
\* 92792 92891: gap of unknown length  
\* 92892 102794: contig of 9903 bp in length  
\* 102795 102894: gap of unknown length  
\* 102895 110866: contig of 7972 bp in length  
\* 110867 110966: gap of unknown length  
\* 110967 117571: contig of 6605 bp in length  
\* 117572 117671: gap of unknown length  
\* 117672 123738: contig of 6067 bp in length  
\* 123739 123838: gap of unknown length  
\* 123839 130583: contig of 6745 bp in length  
\* 130584 130683: gap of unknown length  
\* 130684 141544: contig of 10861 bp in length  
\* 141545 141644: gap of unknown length  
\* 141645 169109: contig of 27465 bp in length  
\* 169110 169209: gap of unknown length  
\* 169210 193446: contig of 24237 bp in length.

FEATURES  
source  
1. .193446  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_lib="RPCT human BAC library 11"  
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN

Query Match 17.6%; Score 27; DB 2; Length 193446;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCGGTGACACTACCTCATCAAA 102  
|||||  
Db 87703 CAGACCGGTGACACTACCTCATCAAA 87729

RESULT 10  
AC022121/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
AC022121  
VERSION AC022121.6 GI:15375145  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 219258)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submissions  
2 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submissions  
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submissions  
JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submissions  
JOURNAL Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced g1:15148108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
STS Content:  
WT-5811 G04974  
WT-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841.  
Location/Qualifiers  
1. .219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

FEATURES  
source  
1. .219258  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 219258;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCGGTGACACTACCTCATCAAA 102  
|||||  
Db 89405 CAGACCGGTGACACTACCTCATCAAA 89379

RESULT 11  
AC092382/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE.  
AC092382  
VERSION AC092382.1 GI:14589571  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVIFERIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 276181)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
REFERENCE 2 (bases 1 to 276181)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 435334  
Center Clone Name: RPCL-11\_75H1  
-----

Summary Statistics  
Consensus quality: 229677 bases at least Q40  
Consensus quality: 256163 bases at least Q30  
Consensus quality: 260799 bases at least Q20  
Estimated insert size: 271820; sum-of-contigs estimation  
Estimated insert size: 271820; agarose-rip estimation  
Quality coverage: 8.62 in Q20 bases; agarose-rip estimation  
Quality coverage: 5.35 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a working draft sequence. It currently  
\* consists of 47 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
1061: contig of 1061 bp in length  
1062  
1161: gap of unknown length  
1162  
2827: contig of 1666 bp in length  
2828  
2927: gap of unknown length  
2928  
4327: contig of 1300 bp in length  
4328  
5618: gap of unknown length  
5619  
5719: gap of unknown length  
5718: contig of 1291 bp in length  
5719  
6983: contig of 1265 bp in length  
6984  
7083: gap of unknown length  
7084  
8423: contig of 1339 bp in length  
8422: gap of unknown length  
8423  
8522: gap of unknown length  
8523  
9771: contig of 1449 bp in length  
9772  
9871: gap of unknown length  
9872  
11198: contig of 1327 bp in length  
11199  
11298: gap of unknown length  
11299  
12414: contig of 1116 bp in length  
12415  
12514: gap of unknown length  
12515  
13806: contig of 1292 bp in length  
13807  
13906: gap of unknown length  
13907  
15360: contig of 1454 bp in length  
15361  
15460: gap of unknown length  
15461  
16667: contig of 1207 bp in length  
16668  
16767: gap of unknown length  
16768  
18390: contig of 1623 bp in length  
18391  
18490: gap of unknown length  
18491  
19541: contig of 1051 bp in length  
19542  
19641: gap of unknown length  
19642  
21156: contig of 1515 bp in length  
21157  
21257: gap of unknown length  
21258  
22448: contig of 1192 bp in length  
22449  
22549: gap of unknown length  
22550  
24499: contig of 1951 bp in length  
24500  
24500: gap of unknown length  
24501  
26384: contig of 1785 bp in length  
26385  
26485: gap of unknown length  
26486  
27883: contig of 1399 bp in length  
27884  
27884: gap of unknown length

\* 27984 30386: contig of 2403 bp in length  
\* 30387 30486: gap of unknown length  
\* 30487 32480: contig of 1994 bp in length  
\* 32481 32580: gap of unknown length  
\* 32581 35441: contig of 2861 bp in length  
\* 35442 35541: gap of unknown length  
\* 35542 37800: contig of 2259 bp in length  
\* 37801 37900: gap of unknown length  
\* 37901 40587: contig of 2687 bp in length  
\* 40588 40687: gap of unknown length  
\* 40688 44159: contig of 3472 bp in length  
\* 44160 44259: gap of unknown length  
\* 44260 46336: contig of 2377 bp in length  
\* 46337 46736: gap of unknown length  
\* 46737 50082: contig of 3346 bp in length  
\* 50083 50182: gap of unknown length  
\* 50183 53988: contig of 3806 bp in length  
\* 53989 54088: gap of unknown length  
\* 54089 56592: contig of 2504 bp in length  
\* 56593 61352: gap of unknown length  
\* 61353 61452: contig of 4660 bp in length  
\* 61453 64234: contig of 2802 bp in length  
\* 64235 64334: gap of unknown length  
\* 64335 68825: contig of 4471 bp in length  
\* 68826 68925: gap of unknown length  
\* 68926 74393: contig of 5468 bp in length  
\* 74394 74493: gap of unknown length  
\* 74494 80561: contig of 5068 bp in length  
\* 80562 80661: gap of unknown length  
\* 80662 87626: contig of 6965 bp in length  
\* 87627 87726: gap of unknown length  
\* 87727 93600: contig of 5874 bp in length  
\* 93601 93700: gap of unknown length  
\* 93701 102024: contig of 8324 bp in length  
\* 102025 102124: gap of unknown length  
\* 102125 109905: contig of 7781 bp in length  
\* 109906 116737: gap of unknown length  
\* 116738 116837: contig of 6732 bp in length  
\* 116838 126897: gap of unknown length  
\* 126898 126997: contig of 9960 bp in length  
\* 126999 135114: gap of unknown length  
\* 135115 135214: contig of 8217 bp in length  
\* 135215 146713: gap of unknown length  
\* 146714 146813: contig of 11499 bp in length  
\* 146814 157894: gap of unknown length  
\* 157895 157994: contig of 11081 bp in length  
\* 157995 171753: gap of unknown length  
\* 171753 171852: contig of 13758 bp in length  
\* 171853 201687: gap of unknown length  
\* 201688 201787: contig of 29835 bp in length  
\* 201788 231269: gap of unknown length  
\* 231269 231369: contig of 29481 bp in length  
\* 231369 276181: gap of unknown length  
\* 276181 44813: contig of 44813 bp in length.

FEATURES  
source 1. -276181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RPCL-11-75H1"

BASP COUNT 86859 a 51769 c 50026 g 82854 t 4673 others  
ORIGIN

Query Match 17.6% Score 27; DB 2; Length 276181;  
Best Local Similarity 100.0%; P-adj. No. 0.00014;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 76 CAGACCCGTCGACACTACCTCATCAAA 102  
Db 213649 CAGACCCGTCGACACTACCTCATCAAA 213623

RESULT 12  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
source Location/Qualifiers  
1.153  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98  
Db 74 TACAGACCCGTCGAGACTACTCAT 98

RESULT 13  
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999  
DEFINITION delta-crystallin enhancer binding protein, complete sequence.  
ACCESSION D14316  
VERSION D14316.1 GI:391639  
KEYWORDS ORF2.  
SOURCE Gallus gallus (library: lambda gt11) 13 day embryo lens cDNA to mRNA, clone JF11.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
REFERENCE 1 (bases 1 to 2292)  
AUTHORS Funahashi,J.  
TITLE Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel., Aging and Cancer, Tohoku Univ., 4-1 Seiryō-machi, Aoba-ku, Sendai 980-77,  
JOURNAL Japan (Tel:022-272-9499, Fax:022-272-3982)  
2 (sites)  
REFERENCE Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.  
AUTHORS Delta-crystallin enhancer binding protein delta EFL is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis  
JOURNAL Development 119 (2), 433-446 (1993)  
MEDLINE 94116444  
REFERENCE 3 (bases 1 to 2292)  
AUTHORS Funahashi,J.  
JOURNAL Unpublished (1994)  
FEATURES  
source Location/Qualifiers  
1.2292  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/tissue.type="lens"  
/clone.lib="lambda gt11"  
/dev\_stage="13 day embryo"  
257.1939  
CDS

/note="beta-gal fusion protein binds to blocks 10 to 3 of HN fragment of delta-crystallin enhancer."  
/codon\_start=1  
/product="ORF2"  
/protein\_id="BA03262.1"  
/db\_xref="GI:391640"  
/translation="MEIKKCNHCYLIKPPDNEFYNKQDALOHLIRSSGKLLDLK LIRREGNRVLIFSQVMNMLDILAEIKYRQPPFORLDSIKELRKQALDHFNAEG SEDPCFLSTRAGGLGINLASADTVIIPSDMNPQDLQAOARHIGOKOVNIRL VIKGSVEEDILERAKKKMYLDHVIORMDTGTYVHTGTSBSTPPEKKELSATLK FGAEEILFKEPGEFEQPEQMDIDELIKRAETRENEPPLVGDLSQKVANFSMD EDDIELEPERNSRWEEILIPESQRRRIIEERKQLELEITMLPRMNCARQJLSPNSE GRRSRERYSGSDSDSITTEKRRKRGPRPTIPRENIKGFSDAETIRFKSYKKEGF LERLDVARDALVDKSENDLRLGELVHNGCIKALKDSSGQERAGRLGKVGPTF RLSGVQVAKVISHHEELAPLHKSTIPSPDEERKRVYIPCHTKAAHFDIDMGKEDDSN LLVGIETAYAGSNQCVNGRSQNLNTEILPDDPDRIPRONSIRPVOTYSLNT"

BASE COUNT 827 a 385 c 346 g 534 t  
ORIGIN

Query Match 16.3%; Score 25; DB 5; Length 2292;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98  
Db 1904 TACAGACCCGTCGAGACTACTCAT 1928

RESULT 14  
LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent WO9639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source Location/Qualifiers  
1.6608  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98  
Db 4123 TACAGACCCGTCGAGACTACTCAT 4147

RESULT 15  
AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
LOCUS AF004397  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the z chromosome protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS chicken.  
SOURCE

ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
JOURNAL A CHD1 gene is 2 chromosome linked in the chicken Gallus domesticus  
MEDLINE Gene 197 (1-2), 225-229 (1997)  
97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
JOURNAL Direct Submission  
TITLE Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source Location/Qualifiers  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-2"  
228..5654  
/gene="CHD-2"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome  
protein  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501846"  
/translation="MNGHSDSEYRNSGESSRSDSDSGASGSGSSGSSGSSGSS  
SOSGSDSESGSSESSESDSRKQVQAPKADSEPKSPSILLAVQNSAV  
LKKQOQKKAASDGSSESDSSSESDSETRKKHKKHDEDMGSSGVSSTGS  
DSEAEKDGKSSCESESDYEPKPKVSKRPPSRIRKPSGKSTGQKROLDSSEEE  
DDDDYDKRSGRQATVAVSYKEEETKTDSDDLLEVCGEDVPQTEDEFEETIEKPD  
SRIRKAGTASTTIVAAECGPNAGPEKSEKELGEIOLYLKWKGMSHHNTMETERT  
LKOQNVGMKRLDNYRKQDETRKMLKNAPEDEVYVYVCOQELTDLHKOYIVERLI  
ASHQKSAQYDPYCKQWGLPYSECSMEDGALAKRFOARIDEYFSRQSKTPEKD  
CVLKKRPREVALKKQPSYIGHSESLDYLQNLNMLASHWCKNSCILLADMGIG  
KTIOTISFLYLFHEHQLYGFLLRYPLSTLSMQREIQTAFQOMNAVYLGDTSSN  
MRTHEWMPQTRKLFNILLTYEILLKDKSFLGNMAFICGVDEARLKNDSILY  
RLIDEPFNRLILGTPLQNSLKEMLSLHFMPEKSSWEDFEHKGKREYVAS  
LAKLEFPLLRKVKQVKSLLPAVEQILRMESALOKOYKWLITRNKALSKSGK  
STSGFLNIMELKCCNHCYLIKPPDNEFYKQALOHILRSSGLILLDLILRLR  
ENGRNVLIFFQWMLDILAEYLYKQPPFQRLDSIKGELRKQALDHFNAGSEDC  
FLSTRAGGGINLASADTVVIFPSDMNPNDLQARARIGOKKOYVIRLYTGS  
VEEDILERRAKKIVLDHVIQMDTGTVLHTGSTSSSTPPNKEELSAILKFGAE  
LEFPEGEDEQEMDIDELIKRAETRENEPGLVGDDELISOYVANSMDDEIE  
LEPRNSRMEELIPESSORRIIEEERKELEIYMLPMRNCARQISNGSEGRSR  
SRVSGSDSITERRRPRKRRPRITPRENTKGSQAEIRREITSYKKPGPIERD  
AVARDAELVNSKSPDLRLGLGVNGCITALKDSSGQPARGLGKVCPTPIRGY  
QVNAKIVISHEPELPIKSTPDPERRRYVICHRAHPIDMGKEDSNILNGI  
VEYQSGMWMKIDPLSLQKILPDDPKRPAKQLOTRADILIKLNDLARKAQ  
RLAGAGSKRRKTRKRRKAKASKIKELIKSDSPQSEKSDDEEDKDEIVSK  
HLHKRTTEKENEKPEPDIGIKKEKEKREKREKSKKIPLDIPVITITISEPVISE  
NEKREKVNESIÖKEKEVEKEKREKREKREKREKREKREKREKREKREKREKREK  
ESSELHOKTISVCKEKEKRVKALQKQDPEKGLSEROLEHTRÖCLIRIGDHITTEL  
KEITNPOLIKOMKMLIFESKLEFEDARKLHAKIKROESQONHONDISNRY  
THVIRNPVDRLEKLTINHDSSRSTSSDRHLQYHDKHDKRQGDATKSDSKRRY  
SAFNSGDKHDMOHTKÖDSRYTSDSKHRLDHRHRSNLEGNLKDSDRGSHDRSH  
SDHRIHSDHSTSETSHKSSRDYRHSQMDHRAKSGSPKSLDÖRSPYGSRPLG  
HRSPEHSDHSTPEHTWSSRKI"

misc\_feature

4341..4604

/gene="CHD-2"

/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t

ORIGIN

Query Match 16.3%; Score 25; DB 5; Length 6872;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TACAGACCCGTCGAGACTACCTCAT 98

Db 4123 TACAGACCCGTCGAGACTACCTCAT 4147

Search completed: August 3, 2002, 01:42:15  
Job time: 40705 sec

.....

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:50:00 ; Search time 949.74 Seconds  
(without adjustment)

276.589 million cell updates/sec

Title: US-08-973-363-2  
Page: 153

Sequence: 1 ATTTCTTCCAGATGATCCTGA.....CTCAGAGACTTTGTGTGCG 153

Scoring table: OLIGO\_NUC

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database: : N\_Geneseq\_032802:\*

1:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1960.DAT *
2:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1961.DAT *
3:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1962.DAT *
4:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1963.DAT *
5:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1964.DAT *
6:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1965.DAT *
7:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1966.DAT *
8:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1967.DAT *
9:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1968.DAT *
10:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1969.DAT *
11:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1970.DAT *
12:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1991.DAT *
13:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1992.DAT *
14:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1993.DAT *
15:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1994.DAT *
16:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1995.DAT *
17:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1996.DAT *
18:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1997.DAT *
19:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1998.DAT *
20:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA1999.DAT *
21:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA2000.DAT *
22:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA2001.DAT *
23:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA2002.DAT *
24:	/SDSI1/gcgdata/hold-genseq/genseqn-emb1/NA2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	153	100.0	153	18	AAI42756	Mouse CHD-1 gene (
2	54	35.3	1311	19	AAV59280	Altered telomere x
3	25	16.3	153	18	AAI42757	Chicken CHD-1A gene
4	22	15.3	6608	18	AAI42751	Chicken CHD-1A gene
5	22	14.4	13884	20	AAI32578	Enterococcus faecalis
6	20	13.1	153	18	AAI42758	Chicken CHD-1 gene
7	19	12.4	803	23	AA581107	DNA encoding novel
8	18	11.8	905	21	AAI51589	Trichodina reseau
9	18	11.8	5168	23	AA581550	DNA encoding novel

[illegible]

PI Griffiths R, Tiwari B;  
XX WPI: 1997-043127/04.  
DR P-PSDB; AAM08146.  
XX  
PT Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determ. and to control sex of progeny  
XX  
PS Claim 8; Fig 3; 76pp; English.  
XX  
CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of  
CC progeny.  
XX  
SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

Query Match 100.0%; Score 153; DB 18; Length 153;  
Best Local Similarity 100.0%; Pred. No. 4.3e-67;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTTCTTCAGATGATCCTGATTAATAAACACAGCAAAACAGTTACGACCAAAACCA 60  
DB 1 attcttcagatgactctgtataaaacacacagcaaaacagttacgacacaaaaccca 60  
OY 61 CAAGCAAAACAGTTTACAGACCCGTCGAGACTTACCTCAATCAACTTACGAGATCTT 120  
DB 61 caagcaaaacagttacagaccgctgcagactactcactcaactcactagcagatcct 120  
OY 121 GCAAAAAGAGGCTCAGACACTTGTGTCGCG 153  
DB 121 gcaaaaagagagctcagagacttgtgtcgcg 153

RESULT 2  
AAV59280  
ID AAV59280 standard; cDNA; 1311 BP.  
XX  
AC AAV59280;  
XX  
DT 14-DEC-1998 (first entry)  
XX  
DE Altered telomere repeat binding factor 1 gene.  
XX  
KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT 1..1311  
FT /\*tag= a  
FT /product= "A-TRF"  
XX  
XX W09836066-A1.  
XX  
XX 20-AUG-1998.  
XX  
XX 13-FEB-1998; 98WO-US02765.  
XX  
XX 04-FEB-1998; 98US-0018628.  
XX 13-FEB-1997; 97US-0800264.  
XX  
XX (UVRQ ) UNIV ROCKEFELLER.  
XX  
PI Bianchi A, De Lange T, Van Steensel B;

XX  
DR WPI: 1998-480769/41.  
DR P-PSDB; AAV59280.  
XX  
PT Nucleic acid encoding altered telomere repeat binding protein and  
XX related vectors - transformants, hetero-dimers and antibodies, used  
XX PT to inhibit shortening of telomerases caused by ageing or disease,  
XX PT also used to extend life of cells in culture  
XX  
PS Claim 14; Page 110-111; 163pp; English.  
XX  
CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
CC inhibit shortening of telomeres associated with ageing (for cosmetic  
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
CC tumours and viral (including human immune deficiency virus) infection.  
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
CC for expression of recombinant proteins or where intended for subsequent  
CC transplant or for testing, eliminating the need for transformation.  
XX  
SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

Query Match 35.3%; Score 54; DB 19; Length 1311;  
Best Local Similarity 100.0%; Pred. No. 1.4e-17;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCGAGACTTACCTCAAACTA 105  
DB 1258 aaaaacacacagcaaaacagttacagaccgctgcagactcactcaactcaacta 1311

RESULT 3  
AAT42757  
ID AAT42757 standard; DNA; 153 BP.  
XX  
AC AAT42757;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Chick CHD-1A gene fragment.  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KW CHD-1A; CHD-W; W chromosome; ss.  
XX  
OS Gallus sp.  
XX  
FH Key Location/Qualifiers  
FT misc\_difference 52..81  
FT /\*tag= a  
FT /note= "bases 52-81 are a repeat of bases 22-51  
FT and are ignored in the translated amino  
FT acid sequence given in Fig 3"  
XX  
XX W09639505-A1.  
XX  
XX 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96WO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX  
XX WPI: 1997-043127/04.  
XX DR P-PSDB; AAM08147.  
XX  
PT Avian chromodomain-helicase-DNA binding genes determine sex in



PT birds - used for sex determ. and to control sex of progeny  
 XX Claim 8; Fig 3; 76pp: English.  
 XX  
 CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757).  
 CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)  
 CC and the great tit CHD-W gene (see also AA142759). Translated amino  
 CC acid sequences of this region are provided in AAM08146-49. The  
 CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;  
 Query Match 16.3%; Score 25; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 0.0049;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 74 TACAGACCGTCGACACTACTCAT 98  
 ||||||||||||||||||  
 Db 74 tacagaccgtgcagactactcat 98  
 RESULT 4  
 AA142751  
 ID AA142751 standard; CDNA; 6608 BP.  
 XX  
 AC AA142751;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A gene.  
 XX  
 KM Bird; sex determination; chromodomain-helicase-DNA binding 1 Avian;  
 KM CHD-W; chromodomain-helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 228..5390  
 FT /\*tag= a  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 1; Fig 5; 76pp: English.  
 XX  
 CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo DNA  
 CC library using a great tit CHD-W sequence (see also AA142755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AA142756-57). It is located on an autosome or Z  
 CC chromosome. Probes based on CHD-W and CHD-1A give a non-rare  
 CC specific signal on hybridisation to genomic DNA of a non-rare  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.  
 XX  
 SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;  
 Query Match 16.3%; Score 25; DB 18; Length 6608;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 74 TACAGACCGTCGACACTACTCAT 98  
 ||||||||||||||||||  
 Db 4123 tacagaccgtgcagactactcat 4147  
 RESULT 5  
 AA13278/C  
 ID AA13278 standard; DNA; 13884 BP.  
 XX  
 AC AA13278;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Enterococcus faecalis genome contig SRQ ID NO:341.  
 XX  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO9650555-A2.  
 XX  
 PD 12-NOV-1998.  
 XX  
 PF 04-MAY-1998; 98MO-US08985.  
 XX  
 PR 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046555.  
 XX  
 PA (HDMA-) HUMAN GENOME SCI INC.  
 XX  
 PT Barash SC, Dillon PJ, Kunsch CA;  
 PT  
 DR WPI; 1999-045171/04.  
 XX  
 XX  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 XX  
 PS Claim 1; Page 1453-1460; 2084pp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC AA132938 to AA13919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based  
 CC system can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 XX  
 SQ Sequence 13884 BP; 4353 A; 2723 C; 2370 G; 4422 T; 16 other;  
 Query Match 14.4%; Score 22; DB 20; Length 13884;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 TGATCCTGATATAAACACCAA 33  
 |||||||  
 DB 13595 TGATCCTGATATAAACACCAA 13574

RESULT 6  
 AAT42758  
 ID AAT42758 standard; DNA; 153 BP.

XX AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KM CHD-1A; CHD-W; W chromosome; ss.  
 XX Gallus sp.

OS

XX Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /tag- a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"

XX MO9639505-A1.

XX 12-DEC-1996.

PD 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI: 1997-043127/04.

XX P-PSDB; AAM08148.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny

PS Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A - Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAM08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.

SO Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 13.1%; Score 20; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 CAGTACAGACCGTCGAGA 89  
 |||||||  
 DB 70 cagttacagaccctgcaga 89

RESULT 7  
 AAS81107  
 ID AAS81107 standard; cDNA; 803 BP.  
 XX

AC AAS81107;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX

DE DNA encoding novel human diagnostic protein #16911.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.  
 XX Dmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG16920.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 1; SEQ ID NO 16911; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 803 BP; 221 A; 195 C; 240 G; 147 T; 0 other;

Query Match 12.4%; Score 19; DB 23; Length 803;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 AACCAAGCAAAACAGTT 44  
 |||||||  
 DB 173 aaccacaagcaaacagtt 191

RESULT 8  
 AAF15189  
 ID AAF15189 standard; cDNA; 905 BP.  
 XX AAF15189;  
 AC

XX 13-MAR-2001 (first entry)  
DE Trichoderma reesei EST SEQ ID NO:7712.  
XX  
XX Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
OS Trichoderma reesei.  
PN WO200056762-A2.  
PD 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07781.  
XX  
XX 22-MAR-1999; 99US-0273623.  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (NOVO ) NOVO NORDISK AS.  
XX  
XX Bertha RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
PI WPI: 2000-594572/56.  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
XX  
XX Claim 89; Page 3111; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.  
XX  
XX Sequence 905 BP; 226 A; 265 C; 185 G; 218 T; 11 other:  
Query Match 11.8%; Score 18; DB 21; Length 905;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 45 ACAGACCAAAAACACCA 62  
DB 291 acagaccataaacacaca 308

ID AAS81550 standard; cDNA; 5168 BP.  
XX  
XX AAS81550;  
AC  
XX 13-FEB-2002 (first entry)  
DE  
XX DNA encoding novel human diagnostic protein #17354.  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Dermanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
XX  
XX P-PSDB; ABE17363.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID No 17354; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 5168 BP; 1428 A; 1397 C; 1194 G; 1149 T; 0 other:  
Query Match 11.8%; Score 18; DB 23; Length 5168;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 5 TTCACAGATGATCCCTGATA 22  
DB 725 ttccagatgattcccgata 742

RESULT 10  
ID AAK83089 standard; DNA; 225 BP.  
ID AAK83089

XX AAK83089;  
AC  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37901.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235583.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246539.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249287.  
PR 17-NOV-2000; 2000US-0249289.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.

```
PR 06-DEC-2000: 2000US-0251479.
PR 06-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251868.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251889.
PR 08-DEC-2000: 2000US-0251989.
PR 08-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM:
XX WPI: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 37901: 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 225 BP: 52 A; 55 C; 55 G; 63 T; 0 other:
XX
XX
XX Query Match 11.1%; Score 17; DB 22; Length 225;
XX Best Local Similarity 100.0%; Pred. No. 49;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 125 AAAGAGGCTCAGAGA 141
XX |||||||||||||||
Db 19 aaagagagctcagaga 35
XX
XX RESULT 11
XX AAK83090
XX ID AAK83090 standard: DNA: 225 BP.
XX
XX AAK83090:
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37902.
XX
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX MO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001: 2001WO-US01354.
XX
XX 31-JAN-2000: 2000US-0179065.
XX
XX 04-FEB-2000: 2000US-0180628.
XX 24-FEB-2000: 2000US-0184664.
XX 02-MAR-2000: 2000US-0186350.
XX 16-MAR-2000: 2000US-0186874.
XX 17-MAR-2000: 2000US-0190076.
XX 18-APR-2000: 2000US-0198123.
XX 19-MAY-2000: 2000US-0205515.
XX 07-JUN-2000: 2000US-0209447.
XX 28-JUN-2000: 2000US-0214886.
XX 30-JUL-2000: 2000US-0215137.
XX 07-JUL-2000: 2000US-0216847.
XX 11-JUL-2000: 2000US-0216889.
XX 11-JUL-2000: 2000US-0217487.
XX 14-JUL-2000: 2000US-0218290.
XX 26-JUL-2000: 2000US-0220963.
XX 26-JUL-2000: 2000US-0220964.
XX 14-AUG-2000: 2000US-0224518.
XX 14-AUG-2000: 2000US-0224519.
XX 14-AUG-2000: 2000US-0225213.
XX 14-AUG-2000: 2000US-0225214.
XX 14-AUG-2000: 2000US-0225266.
XX 14-AUG-2000: 2000US-0225267.
XX 14-AUG-2000: 2000US-0225268.
XX 14-AUG-2000: 2000US-0225270.
XX 14-AUG-2000: 2000US-0225447.
XX 14-AUG-2000: 2000US-0225757.
XX 14-AUG-2000: 2000US-0225758.
XX 14-AUG-2000: 2000US-0225759.
XX 18-AUG-2000: 2000US-0226279.
XX 22-AUG-2000: 2000US-0226681.
XX 22-AUG-2000: 2000US-0226868.
XX 23-AUG-2000: 2000US-0227182.
XX 30-AUG-2000: 2000US-0227009.
XX 01-SEP-2000: 2000US-0228924.
XX 01-SEP-2000: 2000US-0229287.
XX 01-SEP-2000: 2000US-0229343.
XX 01-SEP-2000: 2000US-0229344.
XX 01-SEP-2000: 2000US-0229345.
XX 05-SEP-2000: 2000US-0229509.
XX 05-SEP-2000: 2000US-0229513.
XX 06-SEP-2000: 2000US-0230437.
XX 08-SEP-2000: 2000US-0230438.
XX 08-SEP-2000: 2000US-0231242.
XX 08-SEP-2000: 2000US-0231243.
XX 08-SEP-2000: 2000US-0231244.
XX 08-SEP-2000: 2000US-0231413.
XX 08-SEP-2000: 2000US-0231414.
XX 08-SEP-2000: 2000US-0232080.
XX 08-SEP-2000: 2000US-0232081.
XX 12-SEP-2000: 2000US-0232081.
XX 14-SEP-2000: 2000US-0232397.
XX 14-SEP-2000: 2000US-0232398.
XX 14-SEP-2000: 2000US-0232399.
XX 14-SEP-2000: 2000US-0232400.
XX 14-SEP-2000: 2000US-0232401.
XX 14-SEP-2000: 2000US-0233063.
XX 14-SEP-2000: 2000US-0233064.
XX 14-SEP-2000: 2000US-0233065.
XX 21-SEP-2000: 2000US-0234223.
XX 21-SEP-2000: 2000US-0234224.
XX 25-SEP-2000: 2000US-0234474.
XX 25-SEP-2000: 2000US-0234996.
XX 25-SEP-2000: 2000US-0235484.
XX 27-SEP-2000: 2000US-0235834.
XX 27-SEP-2000: 2000US-0235835.
XX 29-SEP-2000: 2000US-0236327.
XX 29-SEP-2000: 2000US-0236367.
XX 29-SEP-2000: 2000US-0236368.
XX 29-SEP-2000: 2000US-0236369.
XX 29-SEP-2000: 2000US-0236370.
XX 02-OCT-2000: 2000US-0236802.
XX 02-OCT-2000: 2000US-0237037.
```

PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239933.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	06-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251866.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCT INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-483426/52.	
XX		
XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis -	
XX		

PS	Disclosure:	SEQ ID NO 37902;	3071pp +	Sequence Listing;	English.
XX					
CC	AAK54951 to AAK64702	encode the human immune/haematopoietic antigen (I)			
CC	amino acid sequences given in AAM82170 to AAM91921.	(I) have cytostatic			
CC	activity, and can be used in gene therapy and vaccine production.	(I)			
CC	proteins and polynucleotides may be used in the prevention, diagnosis and				
CC	treatment of diseases associated with inappropriate (I) expression. For				
CC	example, they may be used to treat disorders associated with decreased				
CC	expression by rectifying mutations or deletions in a patient's genome				
CC	that affect the activity of (I) by expressing inactive proteins or to				
CC	supplement the patients own production of (I). Additionally, (I)				
CC	polynucleotides may be used to produce the secreted (I), by inserting				
CC	the nucleic acids into a host cell and culturing the cell to express the				
CC	protein. (I) proteins and polynucleotides may be used to prevent,				
CC	diagnose and treat immune/haematopoietic-related diseases, especially				
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703				
CC	to AAK87694 represent human immune/haematopoietic antigen genomic				
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169				
CC	represent sequences used in the exemplification of the present invention.				
XX					
SO	Sequence 225 BP; 52 A; 55 C; 55 G; 63 T; 0 other;				
	Query Match	11.1%;	Score 17;	DB 22;	Length 225;
	Best Local Similarity	100.0%;	Pred. No. 49;		
	Matches 17; Conservative	0;	Mismatches	0;	Indels 0;
					Gaps 0;
QY	125 AAGAGAGGCTCAGACA 141				
Db	19 aaagaagagctcagaga 35				
	RESULT 12				
	AAK90734				
ID	AAK90734 standard; DNA: 225 BP.				
XX					
AC	AAK90734;				
XX					
DT					
XX	05-NOV-2001 (first entry)				
DE					
XX	Human digestive system antigen genomic sequence SEQ ID NO: 4310.				
KW	Human; digestive system antigen; gene therapy; cancer; appendicitis;				
KW	ulcerative colitis; infection; Hirschprung's disease; chronic colitis;				
KW	digestive system disorder; Meckel's diverticulum; ds.				
XX					
OS	Homo sapiens.				
PN					
XX	WO200155314-A2.				
XX					
PD	02-AUG-2001.				
XX					
PF	17-JAN-2001; 2001WO-US01324.				
XX					
PR	31-JAN-2000; 2000US-0179065.				
PR	04-FEB-2000; 2000US-0180628.				
PR	24-FEB-2000; 2000US-0184664.				
PR	02-MAR-2000; 2000US-0186350.				
PR	16-MAR-2000; 2000US-0189874.				
PR	17-MAR-2000; 2000US-0190076.				
PR	18-APR-2000; 2000US-0198123.				
PR	19-MAY-2000; 2000US-0205515.				
PR	07-JUN-2000; 2000US-0209467.				
PR	28-JUN-2000; 2000US-0214886.				
PR	30-JUN-2000; 2000US-0215135.				
PR	07-JUL-2000; 2000US-0216647.				
PR	07-JUL-2000; 2000US-0216880.				
PR	11-JUL-2000; 2000US-0217487.				
PR	11-JUL-2000; 2000US-0217496.				
PR	14-JUL-2000; 2000US-0218290.				
PR	26-JUL-2000; 2000US-0220963.				
PR	26-JUL-2000; 2000US-0220964.				
PR	14-AUG-2000; 2000US-0224518.				

PR 14-AUG-2000: 2000US-0224519.  
PR 14-AUG-2000: 2000US-0225213.  
PR 14-AUG-2000: 2000US-0225214.  
PR 14-AUG-2000: 2000US-0225266.  
PR 14-AUG-2000: 2000US-0225267.  
PR 14-AUG-2000: 2000US-0225268.  
PR 14-AUG-2000: 2000US-0225270.  
PR 14-AUG-2000: 2000US-0225447.  
PR 14-AUG-2000: 2000US-0225757.  
PR 14-AUG-2000: 2000US-0225758.  
PR 14-AUG-2000: 2000US-0225759.  
PR 18-AUG-2000: 2000US-0226279.  
PR 22-AUG-2000: 2000US-0226681.  
PR 22-AUG-2000: 2000US-0226688.  
PR 23-AUG-2000: 2000US-0227182.  
PR 23-AUG-2000: 2000US-0227009.  
PR 30-AUG-2000: 2000US-0228924.  
PR 01-SEP-2000: 2000US-0229287.  
PR 01-SEP-2000: 2000US-0229343.  
PR 01-SEP-2000: 2000US-0229344.  
PR 01-SEP-2000: 2000US-0229345.  
PR 05-SEP-2000: 2000US-0229509.  
PR 05-SEP-2000: 2000US-0229513.  
PR 06-SEP-2000: 2000US-0230437.  
PR 06-SEP-2000: 2000US-0230438.  
PR 08-SEP-2000: 2000US-0231242.  
PR 08-SEP-2000: 2000US-0231243.  
PR 08-SEP-2000: 2000US-0231244.  
PR 08-SEP-2000: 2000US-0231143.  
PR 08-SEP-2000: 2000US-0231144.  
PR 08-SEP-2000: 2000US-0232080.  
PR 08-SEP-2000: 2000US-0232081.  
PR 12-SEP-2000: 2000US-0232397.  
PR 14-SEP-2000: 2000US-0232398.  
PR 14-SEP-2000: 2000US-0232399.  
PR 14-SEP-2000: 2000US-0232400.  
PR 14-SEP-2000: 2000US-0232401.  
PR 14-SEP-2000: 2000US-0233063.  
PR 14-SEP-2000: 2000US-0233064.  
PR 14-SEP-2000: 2000US-0233065.  
PR 21-SEP-2000: 2000US-0234223.  
PR 21-SEP-2000: 2000US-0234274.  
PR 25-SEP-2000: 2000US-0234997.  
PR 25-SEP-2000: 2000US-0234998.  
PR 26-SEP-2000: 2000US-0235484.  
PR 27-SEP-2000: 2000US-0235834.  
PR 27-SEP-2000: 2000US-0235836.  
PR 29-SEP-2000: 2000US-0236327.  
PR 29-SEP-2000: 2000US-0236367.  
PR 29-SEP-2000: 2000US-0236368.  
PR 29-SEP-2000: 2000US-0236369.  
PR 29-SEP-2000: 2000US-0236370.  
PR 02-OCT-2000: 2000US-0236802.  
PR 02-OCT-2000: 2000US-0237037.  
PR 02-OCT-2000: 2000US-0237038.  
PR 02-OCT-2000: 2000US-0237039.  
PR 02-OCT-2000: 2000US-0237040.  
PR 13-OCT-2000: 2000US-0239835.  
PR 13-OCT-2000: 2000US-0239837.  
PR 20-OCT-2000: 2000US-0240960.  
PR 20-OCT-2000: 2000US-0241221.  
PR 20-OCT-2000: 2000US-0241785.  
PR 20-OCT-2000: 2000US-0241786.  
PR 20-OCT-2000: 2000US-0241787.  
PR 20-OCT-2000: 2000US-0241808.  
PR 20-OCT-2000: 2000US-0241809.  
PR 20-OCT-2000: 2000US-0241826.  
PR 01-NOV-2000: 2000US-0244617.  
PR 08-NOV-2000: 2000US-0246474.  
PR 08-NOV-2000: 2000US-0246475.  
PR 08-NOV-2000: 2000US-0246476.  
PR 08-NOV-2000: 2000US-0246477.

PR 08-NOV-2000: 2000US-0246478.  
PR 08-NOV-2000: 2000US-0246523.  
PR 08-NOV-2000: 2000US-0246524.  
PR 08-NOV-2000: 2000US-0246525.  
PR 08-NOV-2000: 2000US-0246526.  
PR 08-NOV-2000: 2000US-0246527.  
PR 08-NOV-2000: 2000US-0246528.  
PR 08-NOV-2000: 2000US-0246532.  
PR 08-NOV-2000: 2000US-0246536.  
PR 08-NOV-2000: 2000US-0246609.  
PR 08-NOV-2000: 2000US-0246610.  
PR 08-NOV-2000: 2000US-0246611.  
PR 08-NOV-2000: 2000US-0246613.  
PR 17-NOV-2000: 2000US-0249207.  
PR 17-NOV-2000: 2000US-0249208.  
PR 17-NOV-2000: 2000US-0249209.  
PR 17-NOV-2000: 2000US-0249210.  
PR 17-NOV-2000: 2000US-0249211.  
PR 17-NOV-2000: 2000US-0249212.  
PR 17-NOV-2000: 2000US-0249213.  
PR 17-NOV-2000: 2000US-0249214.  
PR 17-NOV-2000: 2000US-0249215.  
PR 17-NOV-2000: 2000US-0249216.  
PR 17-NOV-2000: 2000US-0249217.  
PR 17-NOV-2000: 2000US-0249218.  
PR 17-NOV-2000: 2000US-0249244.  
PR 17-NOV-2000: 2000US-0249245.  
PR 17-NOV-2000: 2000US-0249254.  
PR 17-NOV-2000: 2000US-0249255.  
PR 17-NOV-2000: 2000US-0249257.  
PR 17-NOV-2000: 2000US-0249259.  
PR 17-NOV-2000: 2000US-0249300.  
PR 01-DEC-2000: 2000US-0250160.  
PR 01-DEC-2000: 2000US-0250391.  
PR 05-DEC-2000: 2000US-0251030.  
PR 05-DEC-2000: 2000US-0251988.  
PR 05-DEC-2000: 2000US-0251989.  
PR 06-DEC-2000: 2000US-0251479.  
PR 08-DEC-2000: 2000US-0251856.  
PR 08-DEC-2000: 2000US-0251868.  
PR 08-DEC-2000: 2000US-0251869.  
PR 08-DEC-2000: 2000US-0251899.  
PR 08-DEC-2000: 2000US-0251990.  
PR 11-DEC-2000: 2000US-0254097.  
PR 05-JAN-2001: 2001US-0259678.  
  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-502630/55.  
DR  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
XX  
PS Disclosure; SEQ ID NO 4310; 986bp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer. Meckel's diverticulum, bacterial dyspepsia, parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention.  
XX  
SQ Sequence 225 BP; 52 A; 55 C; 55 G; 63 T; 0 other:  
  
Query Match 11.1%; Score 17; DB 22; Length 225;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 125 AAGAGAGGCTCAGACA 141

```
Db      19 aagaagagctcagaga 35
|||||
RESULT 13
AAK90736
ID      AAK90736 standard; DNA; 225 BP.
XX
XX      AAK90736;
AC
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human digestive system antigen genomic sequence SEQ ID NO: 4312.
XX
KW      Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW      ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW      digestive system disorder; Meckel's diverticulum; ds.
XX
OS      Homo sapiens.
XX
XX      WO200155314-A2.
XX
PD      02-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US01324.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0189874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205151.
PR      07-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
PR      22-AUG-2000; 2000US-0226868.
PR      22-AUG-2000; 2000US-0227182.
PR      23-AUG-2000; 2000US-0227189.
PR      30-AUG-2000; 2000US-0228924.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      05-SEP-2000; 2000US-0229309.
PR      05-SEP-2000; 2000US-0229511.
PR      06-SEP-2000; 2000US-0230437.
PR      06-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
```

```
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232397.
PR      14-SEP-2000; 2000US-0232398.
PR      14-SEP-2000; 2000US-0232399.
PR      14-SEP-2000; 2000US-0232400.
PR      14-SEP-2000; 2000US-0232401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      14-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234274.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0234998.
PR      26-SEP-2000; 2000US-0235484.
PR      27-SEP-2000; 2000US-0235834.
PR      27-SEP-2000; 2000US-0235836.
PR      29-SEP-2000; 2000US-0236327.
PR      29-SEP-2000; 2000US-0236367.
PR      29-SEP-2000; 2000US-0236368.
PR      29-SEP-2000; 2000US-0236369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237038.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237040.
PR      13-OCT-2000; 2000US-0239935.
PR      13-OCT-2000; 2000US-0239937.
PR      20-OCT-2000; 2000US-0241221.
PR      20-OCT-2000; 2000US-0241785.
PR      20-OCT-2000; 2000US-0241786.
PR      20-OCT-2000; 2000US-0241787.
PR      20-OCT-2000; 2000US-0241808.
PR      20-OCT-2000; 2000US-0241809.
PR      20-OCT-2000; 2000US-0241826.
PR      01-NOV-2000; 2000US-0244617.
PR      08-NOV-2000; 2000US-0246474.
PR      08-NOV-2000; 2000US-0246475.
PR      08-NOV-2000; 2000US-0246476.
PR      08-NOV-2000; 2000US-0246477.
PR      08-NOV-2000; 2000US-0246478.
PR      08-NOV-2000; 2000US-0246523.
PR      08-NOV-2000; 2000US-0246524.
PR      08-NOV-2000; 2000US-0246525.
PR      08-NOV-2000; 2000US-0246526.
PR      08-NOV-2000; 2000US-0246527.
PR      08-NOV-2000; 2000US-0246528.
PR      08-NOV-2000; 2000US-0246532.
PR      08-NOV-2000; 2000US-0246609.
PR      08-NOV-2000; 2000US-0246610.
PR      08-NOV-2000; 2000US-0246611.
PR      08-NOV-2000; 2000US-0246613.
PR      17-NOV-2000; 2000US-0249207.
PR      17-NOV-2000; 2000US-0249213.
PR      17-NOV-2000; 2000US-0249214.
PR      17-NOV-2000; 2000US-0249215.
PR      17-NOV-2000; 2000US-0249216.
PR      17-NOV-2000; 2000US-0249217.
PR      17-NOV-2000; 2000US-0249218.
PR      17-NOV-2000; 2000US-0249244.
PR      17-NOV-2000; 2000US-0249245.
PR      17-NOV-2000; 2000US-0249264.
PR      17-NOV-2000; 2000US-0249265.
```



PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249298.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 DR WPI; 2001-502630/55.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 XX  
 PS Disclosure; SEQ ID NO 4312; 986bp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, peckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.  
 CC  
 SQ Sequence 225 BP; 52 A; 55 C; 55 G; 63 T; 0 other;  
 XX  
 Query Match 11.1%; Score 17; DB 22; Length 225;  
 Best Local Similarity 100.0%; Pred. NO. 49;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 125 AAAGAGAGGCTCAGAGA 141  
 ||||||||||||||||  
 DB 19 aaagaagagctcagaga 35  
 XX  
 RESULT 14  
 AAC05229  
 ID AAC05229 standard; cDNA; 311 BP.  
 XX  
 AC AAC05229;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 9304.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 990S-0122487.  
 XX  
 PA (GEST ) GENSET.

XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 9304; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 CC  
 SQ Sequence 311 BP; 90 A; 74 C; 63 G; 84 T; 0 other;  
 XX  
 Query Match 11.1%; Score 17; DB 21; Length 311;  
 Best Local Similarity 100.0%; Pred. NO. 48;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 130 GAGGCTCAGAGACTTG 146  
 ||||||||||||||||  
 DB 184 gagctcagagacttg 200  
 XX  
 RESULT 15  
 AAC79872/C  
 ID AAC79872 standard; cDNA; 657 BP.  
 XX  
 AC AAC79872;  
 XX  
 DT 09-FEB-2001 (first entry)  
 XX  
 DE Human secreted protein encoding cDNA for gene 24.  
 XX  
 KW Human; secreted protein; cytosolic; antiarthritic; antiasthmatic;  
 KW immunosuppressive; antiarteriosclerotic; antiinflammatory; nocitropic;  
 KW neuroprotective; antidiabetic; tranquiliser; vulnereary; antibacterial;  
 KW antiproliferic; antiarthritic; antirheumatic; candidant; anti-HIV;  
 KW autoimmune disorder; allergic condition; cardiovascular disorder;  
 KW cancer; neurological disease; tissue repair; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200055176-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000MO-US06057.  
 XX  
 PR 12-MAR-1999; 990S-0124142.  
 XX  
 PR 11-JUN-1999; 990S-0138597.  
 XX  
 PR 03-DEC-1999; 990S-0168666.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Ruben SM, Komatsoulis G;  
 PI  
 XX WPI; 2000-638176/61.  
 DR  
 DR P-PSDB; AABA4853.

XX Novel 49 human secreted proteins useful for diagnosis, prevention and  
PT treatment of disorders including neurological, cell proliferative,  
PT cardiovascular, and autoimmune/inflammatory disorders and microbial  
PT infections -

XX Claim 1a; Page 349; 405pp; English.

XX  
CC This invention describes a novel isolated polypeptide (I) comprising an  
CC amino acid sequence at least 95 % identical to a polypeptide sequence  
CC selected from 49 polypeptides encoded by polynucleotide sequences  
CC included in American Type Culture Collection (ATCC) deposit number  
CC 203917, defined in the specification. The products of the invention have  
CC cytostatic, antiarthritic, antilastmatic, immunosuppressive, nootropic,  
CC antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,  
CC tranquiliser, vulnerary, antibacterial, antiporiatic, antiarrhythmic,  
CC antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)  
CC encoding (I) is useful for preventing, treating or ameliorating a medical  
CC condition and for diagnosing a pathological condition or susceptibility  
CC to the condition. (I) is useful for identifying a binding partner which  
CC affects the activity of the polypeptide and for identifying an activity  
CC in a biological sample. (I), (II) or an antibody (IV) specific to (I) is  
CC also useful for treating or preventing a disease, disorder or condition  
CC associated with aberrant expression of (I). Diseases treated or diagnosed  
CC include immune disorders such as autoimmune diseases, blood protein  
CC disorders, anemia, allergic reactions and conditions such as asthma,  
CC organ rejection or graft-versus-host disease, inflammation, hyper  
CC proliferative disorders, cardiovascular disorders such as arterioarterial  
CC fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ  
CC regeneration, cancer, neovascular glaucoma, diabetic retinopathy,  
CC rheumatoid arthritis, psoriasis, diseases associated with increased  
CC apoptosis that include acquired immunodeficiency syndrome (AIDS),  
CC neurological diseases such as Parkinson's disease, viral, bacterial,  
CC fungal or parasitic diseases. They are also used to repair, replace or  
CC protect tissue damage by congenital defects, to treat trauma, in surgery,  
CC including cosmetic plastic surgery, to treat fibrosis, reperfusion injury  
CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent  
CC skin aging due to sunburn, to change a mammal's mental state or physical  
CC state by influencing biorhythms, cardiac rhythms, depression, memory,  
CC stress and for accelerating wound healing. (I), (II) and/or their agonist  
CC or antagonist are useful as food additives or preservatives to increase  
CC or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamin, mineral or other nutritional components. (I) is  
CC useful for screening therapeutic compounds. (II) is useful in forensic  
CC biology for detecting DNA sequences and as diagnostic probes for  
CC detecting the presence of specific mRNA in a particular cell type.

XX  
SQ Sequence 657 BP; 186 A; 125 C; 203 G; 143 T; 0 other;

Query Match 11.1%; Score 17; DB 21; Length 657;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 133 GCTCAGAGACTTGTGG 149  
|||||  
DB 72 GCTCAGAGACTTGTGG 56

Search completed: August 3, 2002, 01:50:02  
Job time: 32187 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:29 ; Search time 7016.61 Seconds

(without alignments)  
294.307 Million cell updates/sec

Title: US-08-973-363-2

Perfect score: 153

Sequence: 1 ATGTTCAGATGATCTGCA.....CTCAGACACTTGTGCTGCG 153

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em.estlba:\*  
2: em.estlhum:\*  
3: em.estlin:\*  
4: em.estlmu:\*  
5: em.estlov:\*  
6: em.estlpl:\*  
7: em.estlro:\*  
8: em.estlrc:\*  
9: gb.estl1:\*  
10: gb.estl2:\*  
11: gb.hic:\*  
12: gb.gss:\*  
13: em.gss.hum:\*  
14: em.gss.lin:\*  
15: em.gss.pln:\*  
16: em.gss.vrc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	102	66.7	619	BBI55356	BBI55356 BBI55356
2	102	66.7	660	BB461065	BB461065 BB461065
3	51	33.3	438	BB830730	BB830730 BB830730
4	35	22.9	446	BB834922	BB834922 BB834922
5	27	17.6	430	AL601246	AL601246 DKF263137
6	27	17.6	547	AI890775	AI890775 W055111.x
7	27	17.6	686	AW997058	AW997058 OY3-BN004
8	27	17.6	821	BF239967	BF239967 601505170
9	27	17.6	866	AUI25712	AUI25712 AUI25712
10	27	17.6	1028	BB895133	BB895133 601436060
11	26	17.0	337	AW996787	AW996787 OY3-BN004
12	26	15.0	1200	AG131555	AG131555 Pan trogl
13	22	14.4	1312	BB858209	BB858209 1024056D0
14	22	14.4	217	BP147109	BP147109 W07A12.Y
15	19	12.4	404	CNS02ED0	AI1931607 Tetraodon
16	19	12.4	620	AI638930	AI638930 AEMTAD31
17	19	12.4	632	A2846867	A2846867 ZM0147B18

18	19	12.4	654	9	AI623066	AI623066 AEMTAD31
19	19	11.8	932	10	BGI69355	BGI69355 602320903
20	19	11.4	329	10	AA856136	AA856136 W83D08.r
21	18	11.8	380	9	AA148843	AA148843 z009e04.r
22	18	11.8	440	12	AO997521	AO997521 RPT1-23-3
23	18	11.8	448	12	AO024759	AO024759 HS.2184.A
24	18	11.8	451	9	AI65697	AI65697 W83D08.Y
25	18	11.8	486	10	RA0931	RA0931 YF76A12.s1
26	18	11.8	481	9	AW634969	AW634969 B127C11.W
27	18	11.8	491	10	B0202933	B0202933 B0209393
28	18	11.8	495	12	BH570623	BH570623 B06RX18TR
29	18	11.8	507	10	BC370667	BC370667 wv13b06.Y
30	18	11.8	517	12	AO168710	AO168710 HS.3166.A
31	18	11.8	530	9	BB752822	BB752822 BB752822
32	18	11.8	545	10	BI066856	BI066856 BI066856
33	18	11.8	587	12	AO093189	AO093189 RPT1-23-4
34	18	11.8	597	10	AO093189	AO093189 B060189
35	18	11.8	635	10	B043427	B043427 B043427
36	18	11.8	658	9	AL638257	AL638257 AL638257
37	18	11.8	658	9	AW639551	AW639551 B184C09.W
38	18	11.8	700	10	BGI76934	BGI76934 602313611
39	18	11.8	703	12	AG144450	AG144450 Pan trogl
40	18	11.8	709	12	AO545887	AO545887 CITR1-E1
41	18	11.8	810	10	BC666296	BC666296 DRABSA07
42	18	11.8	901	10	BE259673	BE259673 601145816
43	18	11.8	969	12	AG134136	AG134136 Pan trogl
44	18	11.8	993	12	CNS06G3K	AL397206 T3 end of
45	18	11.8	1176	10	BC036467	BC036467 602326413

## ALIGNMENTS

RESULT 1  
BBI55356  
LOCUS  
DEFINITION  
BBI55356 RIKEN full-length enriched, 16 days neonate thymus Mus  
musculus cDNA clone A130024116.3' similar to L10A10 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION  
BBI55356  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

EST  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 619)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

TITLE  
JOURNAL  
COMMENT  
On Jun 29, 2000 this sequence version replaced gi:8811286.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sphiro-cho, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,K., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagl,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Matsuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura



prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the cDNA template. The sequence 5'-GATGCCCCCCCCCCCC 3' cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript Ks(+) after bulk excision from lambda FIC I."

Query Match	66.7%;	Score 102;	DB 9;	Length 660;
Best Local Similarity	100.0%;	Pred. No. 3.1e-42;		
Matches 102; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	52	AAAAACACAGCAAAAAAGTTACAGACCCGTGCAGACACTCATCAACAACTACTAGC	111
	502	AAAAACACAGCAAAAAAGTTACAGACCCGTGCAGACACTCATCAACAACTACTAGC	561
Db	112	AGAGATCTTCGAAAAAGAGAGCTCAGAGACTTGTGTGTGCG	153
QY	562	AGAGATCTTCGAAAAAGAGAGCTCAGAGACTTGTGTGTGCG	603

RESULT	3	
LOCUS	BB830730	
DEFINITION	BB830730	438 bp mRNA linear EST 19-NOV-2001
	BB830730	RIKEN full-length enriched, mammary gland RCB-0527
	JY9-MC(B)	CDNA Mus musculus CDNA clone G93003K04 3', mRNA
	sequence.	
ACCESSION	BB830730	
VERSION	BB830730.1	GI:17008973
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS  
Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi.  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 438)  
Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii,  
Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T.,  
Nakamura, M., Nishi, K., Nomura, K., Nunasaki, R., Okazaki, Y., Okido, T.,  
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Saeki, D., Sato, K.,  
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,  
A., Takehashi, F., Takaku, Akhita, S., Tanaka, T., Tomaru, A., Toya, T.,  
Watanuki, A., Yasunishi, A., Yamatsuyu, M. and Hayashizaki, I.  
Riken Encyclopedia of Mouse Full-Length cDNAs (Akimura, T., et al.  
2001)  
JOURNAL  
Unpublished (2001)  
CONTACT  
Contact: Yoshihide Hayashizaki

Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suheiho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: <http://genome-gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Waghi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watabiki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers

1..438

FEATURES

SOURCE

```

/clone.lib="Riken full-length enriched, mammary gland
RCB-0527 Jyg-MC(B) CDNA"
/clisue_type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
BASE COUNT      138 a      96 c      108 g      96 t
ORIGIN

```

Query Match	33.38;	Score 51;	DB 9;	Length 438;
Best Local Similarity	100.0%;	Pred. No. 7.1e-16;		
Matches 51; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

2y 1 ATCTTCCAGATGATCTGTATAAAAAACCAACCAAGCAAAACAGTTACAGACC 51  
 |||||  
 300 ATCTTCCAGATGATCTGTATAAAAAACCAACCAAGCAAAACAGTTACAGACC 350  
 |||||

RESULT	4				
LOCUS	BB834922				
DEFINITION	BB834922	446 bp	mRNA	linear	EST 19-NOV-2001
	BB834922	Riken full-length enriched,	mammary gland	RCB-0527	
	JY9-NC(B)	CDNA	Mus musculus	CDNA clone G930033721	3', mRNA
	sequence.				
ACCESSION	BB834922				
VERSION	BB834922.1				
KEYWORDS	GI:17013165				
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia: Metazoa: Chordata: Craniata: Vertebrata, Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
(Bases 1 to 446)

Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanaagaki, T., Hayashino, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakanuma, M., Nishi, K., Nomura, K., Nunashi, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shirai, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takemashi, F., Takaku, Akahira, S., Tanaka, T., Tomaru, A., Toyo, T., Watanuki, A., Yasunishi, A., Yamamoto, K. and Hayashizaki, T.  
Riken Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki

JOURNAL  
COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matsubuchi, M., Yonekita, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multipicapillary sequencer. *Genome Res.*  
10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

FEATURES  
source  
Location/Qualifiers  
1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033j21"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT  
139 a 99 c 108 g 100 t

Query Match 22.9%; Score 35; DB 9; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTCTCCAGATGATCCTGATTAACACACACAGC 35  
|||||  
Db 309 ATTCTCCAGATGATCCTGATTAACACACAGC 343

RESULT 5  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFP313J1040.F1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 430)  
Bloeker,H., Boecker,M., Brandt,P., Mewes,W., Weill,B. and Wiemann  
S.

REFERENCE 1 (bases 1 to 430)  
AUTHORS Bloeker,H., Boecker,M., Brandt,P., Mewes,W., Weill,B. and Wiemann  
S.  
TITLE EST (Bloeker,H., Boecker,M., Brandt,P., Mewes,W., Weill,B. and  
Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloecker H  
MIPS  
Am Kiopterspitze 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).  
Location/Qualifiers  
1..430

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFP313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTRipleX2; Site\_1: SfiI; Site\_2: SfiI;B;  
cDNA-collection"

BASE COUNT  
163 a 81 c 86 g 100 t

Query Match 17.6%; Score 27; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 263 CAGACCCGTCAGACTACCTCATCAAA 289

RESULT 6  
LOCUS AI890775/c 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wm95f11.x1 NCI CGAP UT2 Homo sapiens cDNA clone IMAGE:2443725 3'  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; mRNA sequence.  
AI890775  
VERSION AI890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert length: 1924 Std Error: 0.00  
Seq primer: -400p from Glibco  
High quality sequence stop: 418.  
Location/Qualifiers  
1..547

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone\_lib="NCI-CGAP-UT2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 Kb. Life Technologies catalog #:  
11539-012"

BASE COUNT  
114 a 118 c 85 g 230 t

Query Match 17.6%; Score 27; DB 9; Length 547;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 333 CAGACCCGTCAGACTACCTCATCAAA 307

RESULT 7  
LOCUS AM997058/c 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM997058

VERSION AM97058.1 GI:8257292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 686)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
/M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-QV3-BN0047-150  
400-152-c03&t3=2000-04-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 678.  
Location/Qualifiers  
1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="BN0047"  
/dev\_stage="Adult"  
/note="Organ: Breast; normal; Vector: puc18; Site: 1: Smal;  
Site: 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 165 a 154 c 126 g 241 t  
ORIGIN  
Query Match 17.6%; Score 27; DB 9; Length 686;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 76 CAGACCCGTCGACGACTACCTCATCAAA 102  
DB 83 CAGACCCGTCGACGACTACCTCATCAAA 57  
RESULT 8  
LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION 60190510F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
mRNA sequence.  
ACCESSION BF239967  
VERSION BF239967.1 GI:11153890  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 821)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LCM1033 row: k column: 18  
High quality sequence stop: 562.  
Location/Qualifiers  
1..821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="IMAGE:4133129"  
/clone\_1lb="NIH\_MGC\_54"  
/tissue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: bone marrow; Vector: pONR-LIB (Clontech);  
Site: 1: SfiI (ggccgctcgcc); Site: 2: SfiI (ggccatctggcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATCTAGAGCGGAGCGCGCGCATG-3' and 3' adaptor  
sequence: 5'-CAGGCGCATATGAGCC-3' (where B = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."  
BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN  
Query Match 17.6%; Score 27; DB 10; Length 821;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 76 CAGACCCGTCGACGACTACCTCATCAAA 102  
DB 33 CAGACCCGTCGACGACTACCTCATCAAA 59  
RESULT 9  
LOCUS AUI25712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AUI25712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AUI25712  
VERSION AUI25712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,  
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and  
Isogai, T.  
HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,  
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki  
, Y., Sugano, S., Isogai, T.)  
Unpublished (2000)  
JOURNAL Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1537-3 Yana, Kisarazu, Chiba 297-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5' - 3' - end one pass sequencing. Helix  
Research Institute, cDNA library construction. Department of

Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source  
1. .866  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_11b="NT2RM4002061"  
/clone\_11b="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 866;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTGACAGTACCTCATCAAA 102  
|||||  
Db 495 CAGACCCGTGACAGTACCTCATCAAA 521

RESULT 10  
BE895133 1028 bp mRNA linear EST 20-OCT-2000  
LOCUS 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
DEFINITION mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS NIH-MGC http://imgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM9753 row: h column: 16  
High quality sequence stop: 488.  
Location/Qualifiers

FEATURES  
source  
1. .1028  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_11b="IMAGE:3921087"  
/clone\_11b="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Skin; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

Query Match 17.6%; Score 27; DB 10; Length 1028;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTGACAGTACCTCATCAAA 102

Db 176 CAGACCCGTGACAGTACCTCATCAAA 202  
|||||

RESULT 11  
AW996787 337 bp mRNA linear EST 05-JUN-2000  
LOCUS QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW996787  
VERSION AW996787.1 GI:8257021  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Rui Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-BN0047-230  
200-102-d03&ct3=2000-02-23&tl=1)  
Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 337.  
Location/Qualifiers

FEATURES  
source  
1. .337  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_11b="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 83 a 59 c 72 g 123 t  
ORIGIN

Query Match 17.0%; Score 26; DB 9; Length 337;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 GATTAATAACACACACCAACAGT 44  
|||||  
Db 87 GATTAATAACACACACCAACAGT 62

RESULT 12  
AG131555 1200 bp DNA linear GSS 04-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-143K14.F, genomic survey sequence.  
DEFINITION  
ACCESSION AG131555  
VERSION AG131555.1 GI:16661233



KEYWORDS GSS: GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_11b:PTB Chimpanzee Male  
BAC library clone:PTB-143K14.F.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.  
REFERENCE 1 (sites)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Toki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
2 (bases 1 to 1200)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Toki, Y., Watanabe, H. and Sakaki, Y.  
DIRECT SUBMISSION  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpes@gsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the Rsd process and may have higher chance of  
clone tracking errors.  
COMMENT PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R.site 1 : SacI  
R.site 2 : SacI.  
Location/Qualifiers  
1. 1200  
FEATURES  
source  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-143K14.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_11b="PTB Chimpanzee Male BAC Library"  
BASE COUNT 526 a 371 c 199 g 39 t 65 others  
ORIGIN  
Query Match 15.0%; Score 23; DB 12; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 50 CCAAAAACCAACGCAACAG 72  
|||||  
Db 254 CCAAAAACCAACGCAACAG 276  
RESULT 13  
LOCUS B6858209 1312 bp mRNA linear EST 29-MAY-2001  
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION B6858209  
VERSION B6858209.1 GI:14239393  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadae.  
REFERENCE 1 (bases 1 to 1312)  
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,  
McMurtrei, J.P., Sillflow, C., Stern, D., Surzycki, R.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants: project phase 2  
JOURNAL Unpublished (2000)  
CONTACT: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000

FEATURES  
source  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chausere@duke.edu.  
Location/Qualifiers  
1. 1312  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_11b="C. reinhardtii CC-1690, normalized, lambda zap  
IT"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McMurtrei, combines cDNAs from CC-1690 cells grown to  
mid-log phase in RAP (acetate-containing) medium in the  
light "RAP medium in the dark" HS (minimal) medium in  
ambient levels of CO2 and HS medium bubbled with 5% CO2.  
PolyA mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with Exsist 4 (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."  
BASE COUNT 529 a 368 c 380 g 15 t 20 others  
ORIGIN  
Query Match 14.4%; Score 22; DB 10; Length 1312;  
Best Local Similarity 100.0%; Pred. No. 0.71;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 51 CAAAACCAACGCAACAG 72  
|||||  
Db 319 CAAAACCAACGCAACAG 340  
RESULT 14  
LOCUS BF147109/c 217 bp mRNA linear EST 29-DEC-2000  
DEFINITION uyo7a12.y1 McCarrey Eddy spermatoocytes Mus musculus cDNA clone  
IMAGE:3657310 5' similar to SW:TCP2.MOUSE P11983 T-COMPLEX PROTEIN  
1, ALPHA SUBUNIT B ; , mRNA sequence.  
ACCESSION BF147109  
VERSION BF147109.1 GI:11028504  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 217)  
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, F., Person,  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Riller,  
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R. and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1419614  
Seq primer: Primer name ambiguous  
High quality sequence stop: 74.  
FEATURES  
source  
1. 217  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"

```
/clone="IMAGE:3657310"
/clone_lib="McCarrey Eddy spermatocytes"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pBluescript SK+ (Stratagene
); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
[5'-(GA)10-ACTAGCTCTGAGTTTCTTTT-3'] and directionally
cloned using 5' linkers 5'-AATCGGCACGACG-3' and
5'-CTCGTCGCG-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."
```

BASE COUNT 46 a 38 c 60 g 73 t

ORIGIN

Query Match 12.4%; Score 19; DB 10; Length 217;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 TAGCAGAGATCTTGCAAAA 126  
|||||  
Db 193 TAGCAGAGATCTTGCAAAA 175

## RESULT 15

CNS02EDQ 404 bp DNA linear GSS 13-MAY-2000  
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Orl end of clone

DEFINITION 262M09 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL193607.1 GI:7831713  
VERSION GSS: genome survey sequence.  
KEYWORDS Tetraodon nigroviridis.  
SOURCE Tetraodon nigroviridis  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorphna; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 404)  
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 404)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 404)

AUTHORS Genoscope.

JOURNAL Direct Submission  
TITL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
Location/Qualifiers

```
source 1..404
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="262M09"
/clone_lib="G"
/notes="Genoscope sequence ID : C0AG262AG05SP1-end :
PUC-Orl"
```

BASE COUNT 103 a 104 c 102 g 84 t 11 others

ORIGIN

Query Match 12.4%; Score 19; DB 12; Length 404;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AAAAACCACAGCAAAACA 41  
|||||  
Db 352 AAAAACCACAGCAAAACA 370

Search completed: August 2, 2002, 22:41:33  
Job time: 30148 sec

Mon Aug 5 11:51:48 2002

us-08-973-363-2.011.rst

Page 9



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:42:15 ; Search time 10310.5 Seconds  
(without alignments) 310.532 Million cell updates/sec

Title: US-08-973-363-3

Sequence: 1 ATTTCCTGATGATCCAGA.....CACAAAGCTTGCTGTGCA 153

Scoring table:

Gapop 60.0 , Gapext 60.0

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ha.\*  
2: gb\_hng.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ha.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_of.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hng\_hum.\*  
31: em\_hng\_inv.\*  
32: em\_hng\_other.\*  
33: em\_hng\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	153	100.0	153	6	A58684	A58684 Sequence 3
2	103	67.3	6608	6	A58691	A58691 Sequence 10
3	103	67.3	6672	5	AF004397	AF004397 Gallus ga
4	64	41.8	2292	5	D14316	D14316 delta-eryt
5	47	30.7	153	6	A58685	A58685 Sequence 4
6	47	30.7	1316	6	A58696	A58696 Sequence 15
7	38	24.8	153	6	A58686	A58686 Sequence 5
8	25	16.3	153	6	A58683	A58683 Sequence 2
9	25	16.3	5349	10	AR029026	AR029026 Sequence
10	25	16.3	5947	10	MUSCHD1X	MUSCHD1X
11	23	15.0	5947	9	AF006513	AF006513 Mouse DNA-b
12	23	15.0	101220	9	AC092372	AC092372 Homo sapi
13	23	15.0	134365	9	AC012624	AC012624 Homo sapi
14	23	15.0	143079	9	AC021449	AC021449 Homo sapi
15	23	15.0	145659	2	AC008531	AC008531 Homo sapi
16	23	15.0	193446	2	AC091946	AC091946 Homo sapi
17	23	15.0	219258	2	AC022121	AC022121 Homo sapi
18	21	13.7	92881	2	AC092382	AC092382 Homo sapi
19	21	13.7	92881	2	AC108265	AC108265 Rattus no
20	20	13.1	2754	5	AF181824	AF181824 Aegolius
21	20	13.1	2754	5	AF181827	AF181827 Nymphicus
22	20	13.1	2943	5	AF181826	AF181826 Gallus ga
23	20	13.1	147080	9	AL391666	AL391666 Human DNA
24	20	13.1	147891	2	AC019074	AC019074 Homo sapi
25	20	13.1	183892	2	AC098997	AC098997 Rattus no
26	20	13.1	185204	2	AL446026	AL446026 Homo sapi
27	20	13.1	302097	2	AL446026	AL446026 Homo sapi
28	19	12.4	3347	3	DM414676	DM414676 Drosophila
29	19	12.4	3362	3	CEC31810	CEC31810 Caenorhabdit
30	19	12.4	30636	3	AC101875	AC101875 Mus muscu
31	19	12.4	51832	2	AC101875	AC101875 Mus muscu
32	19	12.4	51832	2	LMF1CHR18.01	LMF1CHR18.01
33	19	12.4	110000	2	AL353133	AL353133 Human DNA
34	19	12.4	123527	9	AL353133	AL353133 Human DNA
35	19	12.4	127381	9	AL356268	AL356268 Homo sapi
36	19	12.4	129473	9	AC040158	AC040158 Homo sapi
37	19	12.4	130138	9	AB053170	AB053170 Homo sapi
38	19	12.4	136500	9	AC097067	AC097067 Homo sapi
39	19	12.4	143900	2	AC092137	AC092137 Homo sapi
40	19	12.4	157600	2	AC092539	AC092539 Homo sapi
41	19	12.4	165536	9	AC093600	AC093600 Homo sapi
42	19	12.4	169514	9	AC064869	AC064869 Homo sapi
43	19	12.4	179619	2	AC018596	AC018596 Homo sapi
44	19	12.4	182199	2	AC021805	AC021805 Homo sapi
45	19	12.4	182500	2	AC026988	AC026988 Homo sapi

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT	FEATURES	BASE COUNT	ORIGIN
1	A58684	Sequence 3 from Patent WO9639505.	A58684	A58684.1	GI:3714247	unclassified.	unclassified.	unclassified.	unclassified.	unclassified.	unclassified.	unclassified.	58 a	40 c
													31 g	24 t

Patent: WO 9639505-A 3 12-DEC-1996;  
 ILS INNOVATION (GB) 5906996 961224.  
 Other publication AU 9606996 961224.  
 Location/Qualifiers  
 1..153  
 /organism="unclassified"  
 /db\_xref="taxon:32644"

Query Match	100.0%	Score 153:	DB 6:	Length 153:
Best Local Similarity	100.0%	Pred. No. 4.3e-49:		
Matches 153:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	1	ATTTCCTGATGATTCAGACCAAGAAACCCGAGGCAAGACAGCTACAGACCAAGAAACCC	60	
Db	1	ATTTCACCTGATGATTCAGACCAAGAAACCCGAGGCAAGACAGCTACAGACCAAGAAACCC	60	
OY	61	CAGGCAAGACAGCTACAGACCCGCTGACAGCTCCATTAATTCGTAAATAAGACCTT	120	
Db	61	CAGGCAAGACAGCTACAGACCCGCTGACAGCTCCATTAATTCGTAAATAAGACCTT	120	
OY	121	GCAAGAAAGGAAGACCAAAAGGCTTCTGTGCA	153	
Db	121	GCAAGAAAGGAAGACCAAAAGGCTTCTGTGCA	153	
RESULT	2			
LOCUS	A58691	6608 bp	DNA	linear
DEFINITION	Sequence 10 from Patent WO9639505.			
ACCESSION	A58691			
VERSION	A58691.1	GI:3714250		
KEYWORDS				
SOURCE	unidentified.			
ORGANISM	unidentified.			
REFERENCE	unclassified.			
AUTHORS	1 (bases 1 to 6608)			
TITLE	Griffiths, R. and Tiwari, B.			
JOURNAL	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS			
COMMENT	Patent: WO 9639505-A 10 12-DEC-1996;			
FEATURES	ISIS INNOVATION (GB)			
source	Other publication AU 5906996 961224.			
BASE COUNT	2289 a 1207 c 1459 g 1653 t			
ORIGIN	1. 6608			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
Query Match	67.3%	Score 103:	DB 6:	Length 6608:
Best Local Similarity	100.0%	Pred. NO. 5.3e-49:		
Matches 103:	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	51	CAGAAACCCCGAGCAAGACAGCTACAGACCCGCTGACAGCTACCTTAATTAATCTGAA	110	
Db	4100	CAGAAACCCCGAGCAAGACAGCTACAGACCCGCTGACAGCTACCTTAATTAATCTGAA	4159	
OY	111	TAAAGACCTTCGACAGAAAGGAAGACAAAGGCTTCTGTGCA	153	
Db	4160	TAAAGACCTTCGACAGAAAGGAAGACAAAGGCTTCTGTGCA	4202	
RESULT	3			
LOCUS	AF004397	6672 bp	mRNA	linear
DEFINITION	Gallus gallus chromo-helicase-DNA-binding on the z chromosome			
VERSION	AF004397			
KEYWORDS	protein, variant with hydrophilic domain, (CHD-z) mRNA, complete cds.			
SOURCE	AF004397.1	GI:2501845		
ORGANISM	chicken.			
AUTHORS	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
	Phasianinae; Gallus.			
	1 (bases 1 to 6672)			
	Griffiths, R. and Korn, R.M.			

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	JOURNAL	FEATURES
A CHD1 gene is z chromosome linked in the chicken gallus domesticus	Gene 197 (1-2), 225-229 (1997)	2 (bases 1 to 6872)	Griffiths, R. and Korn, R.M.	Submitted (16-MAY-1997)	Zoology, Molecular Lab, Glasgow University, Glasgow G12 8QQ, UK	location/Qualifiers
source		1..6872	/organism="Gallus gallus"			
gene		/db_xref="taxon:9031"				
CDS		1..6872	/gene="CHD-Z"			
		228..5654	/gene="CHD-Z"			
			/function="role in chromatin architecture"			
			/note="CHD protein with hydrophilic domain"			
			/codon_start=1			
			/product="chromo-helicase-DNA-binding on the z chromosome protein"			
			/protein_id="AAC60282.1"			
			/db_xref="GI:2501846"			
			/translation="MNGSDSESYRNSGSSRSDDSGASGSGSGSSGSSDGS			
			SQSSSSSESGSESGSESSEDSFTRKKQYQAAPPADSEFWKSPSLIAVQRSAV			
			LKROOQOKAASDVGSEEDSSSESDADSSSTTKKKHKDEDMQSSGVSVTGS			
			DSEAEEDGKSCSESESDYKKAETVKKRPPRIKPSGKSTGQKROLDSSEEE			
			DDDDYDGRSGRQATVYVYKKAETKTSDLLCYGDSQVQTEDEDEETLEKPYD			
			SRIGKATGASTTIYAVADQNGAEFEKSKELGEIYQYLKKMGSHIMNTDEET			
			LKQNGKMGKLDNYKKKDDQETKRLKNASPEDEYENCOYLTDLDDHKQYVERTI			
			AHSQKSAAGIPDYCKWMQGLPYSECSMEDGALLAKFKQARIDETFSNKQKTPK			
			CKYLKQRPREFVALKQPSYIGGHESELRLRYQLNGLMWLASHWCKSGSLIADMGIG			
			KTIQITISFLNYLFHEHQLGPEFLRVLPLSTLSMORLQIWAQPMANAVYLIGITSN			
			MIRHEMWHNPQTKRLKFNILITLYEILKKSFGFQANMAFIDEAHRKLDISLY			
			RTLDKFSNRLITGTPLQNSLKEILSLAFIEPEKFSSEDEPEHGRKREGVAS			
			LKHELEPELLRLRYKKDYKSLPAKVEQILMEMSALOKQYKKMLTPRYKALSKGSG			
			STGSELINMEILKCCCHNCLIKRPDQNETRYNQOALQHLIRSSGKILINDKLILR			
			ERGRVNLFSQVNRMDILIAEYLKTRQFPQRDLGSLTGELRKQALDHFNAEGSEDF			
			FLDSTRAGGLGINLASADTVYFDSDNPONDLOAQRAHRIQOKQVNTYLRVTKGS			
			VEEDILERRAKKQMLDLVLQRMDDTQGTLYLHGTSPTSSPNKEKLSAILFGABEE			
			LFKEPEGEDEPOEMOIDEILKRAETRENPGLTVGDELSLOFYKVFNMEDDIE			
			LEPRNRNMEELIPESORRIIEEEROKLEETLYMLPRMRNAKQISTFNGSEGRSR			
			SRRSQSDSDSTIERKKRPKKRGPRPTIPRENIQSGFQALIRPTIKYKKRGPLEERD			
			AVARDALDYKSETDLRISGEILYNGCTIKALONSSQEPKAGGRGLKVKQPTRISSV			
			QVANKVLTSHEELIAPLHKSLPSDPEERKRVIPCHTKAHHFIDWKEEDSNLVTI			
			YEGYGSWEMIKMDPDLSTLQKLLIDPDRPKPAKQLOTDADVYLKILNLDLAKRKQ			
			RLGAGNSRRKRTFRNKKRNKASKIKKEIKSDSPQSEDEDEEDDKMDLVSQYK			
			HLHKIKTEKENEEKRPDIDIGIKKEAEKRETEKENKRLKREKKKEKKELKED			
			NKREKRNKVESIQKEKEVEKVENKSNKESKSLIPLDVPHITATASEPVISLE			
			ESEELHOKTFSQEKERMRPYKALIKOIDREKIGSEEDQLEHTRQCLIKQDHTETL			
			KEYTNPQIKQWRKNMLIPTYKTEFPARKLIKTIKRIKKROESQOILNQONISSYVL			
			THVIRNPDEVERLEKTTNHDSSRDSYSDSNHLSQYDHHNDROGDYKKSDDSKRRY			
			SAPSNGDHRSDWHYKQDSRYSDSKRIKLDHDSNRHNSLENLDSGSHSDHRS			
			SDHSHSDHSTSEYSHHKSRSQYRHSWDMQMDHRSAGSDPRSDLRDSRSPGLG			
			HRSPFHSDDKSPREHTWSRKT"			
			4341..4604			
misc_feature			/gene="CHD-Z"			
			/note="Short insert found in longer variant mRNA of CHD-Z"			
BASE COUNT	2446 a	1233 c	1520 g	1693 t		
ORIGIN						
Query Match		67.3%	Score 103;	DB 5;	Length 6872;	
Best Local Similarity	100.0%	Pred. No. 5.2e-49;				
Matches 103;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	51 CAAGAAACCCAGGCAAGCAGCTACAGACCCGTCACAGCTACCTCATTAATTAAGTGA	110				
	AA	111				
DB	4100 CAAGAAACCCAGGCAAGCAGCTACAGACCCGTCACAGCTACCTCATTAATTAAGTGA	4159				
QY	111 TTAAGACCTTGCAAGAAAGAAAGCAAAAGCCTTGCTGCTGCA	153				
	AA	154				
DB	4160 TTAAGACCTTGCAAGAAAGAAAGCAAAAGCCTTGCTGCTGCA	4202				

```
RESULT 4
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999
DEFINITION delta-crystallin enhancer binding protein, complete sequence.
ACCESSION D14316
VERSION D14316.1 GI:391639
KEYWORDS ORF2
SOURCE Gallus gallus (Library: lambda gtl1) 13 day embryo lens cDNA to
mRNA, clone JF11.
ORGANISM Gallus gallus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2292)
AUTHORS Funahashi, J.
JOURNAL Direct Submission
Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel. Aging and
Cancer, Tohoku Univ.; 4-1 Seiryomachi, Aoba-ku, Sendai 980-77,
Japan (Tel:022-272-9499, Fax:022-272-3982)
2 (sites)
REFERENCE Funahashi, J., Sekido, R., Mura, K., Kamachi, Y. and Kondoh, H.
AUTHORS Delta-crystallin enhancer binding protein delta EPI is a zinc
finger-homeodomain protein implicated in postgastrulation
embryogenesis
JOURNAL Development 119 (2), 433-446 (1993)
MEDLINE 9411644
REFERENCE 3 (bases 1 to 2292)
AUTHORS Funahashi, J.
JOURNAL Unpublished (1994)
FEATURES
source
location/Qualifiers
1..2292
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="lens"
/cclone_lib="lambda gtl1"
/dev_stage="13 day embryo"
257..1939
/note="beta-gal fusion protein binds to blocks 10 to 3 of
HN fragment of delta-crystallin enhancer."
/codon_start=1
/product="ORF2"
/protein_id="BA03262.1"
/db_xref="GI:391640"
/translation="MELKRCNHCYLKPPDNEFYNNQALQHLIRSSGKLLLDKL
LIRKREGRNRYLLFSQVRMDITAEYLRKRPFRQDLSIKELRKQALDHPNAG
SDPCFLSTNAGGLGINLASADTVIFSDNMPQNDQAARAHRIQKQKVNAYRL
VKGSEVEDILERAKKRWLDLVYQRMDTGKTVLHGTSPSSSTPFNKEELSAIK
EADLEPERGEDEQPEMDIDELKRAETRENEPRTYVGLDLSQFVANFSNMD
EODILEPERSRNMEELIPESQRIIEERQKLEELTYMLPRMRCAKOISFNGSE
GRRSRKRYSGSDSISITERKRPKRGKRPRTIPRENKGFSDAIEIRFIKSYKKGPF
LRRDAVARDAELVDKSETDLRRLGELVHNGCIKALKNGSGORAGRLGVKGPTE
RISGVYNAKIVISHEELAPLHKISPSDEEKRRVIPCRTKAHFDIDWCKEDDSN
LIVGIYEYAVGSMQCVNCSRLQNTLELDDPDRTPRONSRYRVTTSINY"
BASE COUNT 827 a 385 c 546 g 534 t
ORIGIN

Query Match 41.8%; Score 64; DB 5; Length 2292;
Best Local Similarity 100.0%; Pred. No. 3.1e-26;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 CAGCTACGAGACCGTGCACATCTATTAAATTCAGATAAGACCTTCAGAGAAG 129
|||||
DB 1900 CAGCTACGAGACCGTGCACATCTATTAAATTCAGATAAGACCTTCAGAGAAG 1959

OY 130 GAAG 133
|||||
DB 1960 GAAG 1963

RESULT 5

RESULT 6
LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 15 from Patent W09639505.
ACCESSION A58696
VERSION A58696.1 GI:3714253
KEYWORDS unclassified.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1316)
AUTHORS Griffiths, R. and Tlwar, B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
BIRDS
JOURNAL Patent: WO 9639505-A 15 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES
source
location/Qualifiers
1..1316
/organism="unclassified"
/db_xref="taxon:32644"
BASE COUNT 493 a 205 c 308 g 304 t 6 others
ORIGIN

Query Match 30.7%; Score 47; DB 6; Length 1316;
Best Local Similarity 100.0%; Pred. No. 2.6e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTTACTGATAAAGACCTTCGAGAAAGAGACACA 137
|||||
DB 1198 TACCTCATTAATTTACTGATAAAGACCTTCGAGAAAGAGACACA 1244

RESULT 7
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 5 from Patent W09639505.
ACCESSION A58686
VERSION A58686.1 GI:3714249
```

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
/organism="unidentified"  
BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

Query Match 24.8%; Score 38; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 5.2e-11;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTACGATTAAGACCTTGCAAGAA 128  
Db 91 TACCTCATTAATTACGATTAAGACCTTGCAAGAA 128

RESULT 8  
A58683  
LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246  
KEYWORDS unidentified.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R, and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
/organism="unidentified"  
BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCGTGCAGACTACTCAT 98  
Db 74 TACAGACCGTGCAGACTACTCAT 98

RESULT 9  
AR029026  
LOCUS AR029026 1311 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5859183.  
ACCESSION AR029026  
VERSION AR029026.1 GI:5940999  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unidentified.  
REFERENCE 1 (bases 1 to 1311)

AUTHORS de Lange,T., Steensel,Bvan, and Bianchi,A.  
TITLE Altered telomere repeat binding factor  
JOURNAL Patent: US 5859183-A 3 12-JAN-1999;  
FEATURES Location/Qualifiers  
source 1..1311  
/organism="unknown"  
BASE COUNT 441 a 257 c 316 g 297 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 1311;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCGTGCAGACTACTCAT 98  
Db 1280 TACAGACCGTGCAGACTACTCAT 1304

RESULT 10  
MUSCHD1X  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410.1  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 5349)  
AUTHORS Delmas,Y., Stokes,D.G. and Perry,R.P.  
TITLE A mammalian DNA-binding protein that contains a chromodomain and an SRF2/SWI2-like helixase domain  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
MEDLINE 93211972  
REFERENCE 2 (bases 1 to 5349)  
AUTHORS Perry,R.P.  
TITLE Direct Submission  
JOURNAL Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES Location/Qualifiers  
source 1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AA08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDSESVRNGSGESSQSGDCCGASGSGSGSSGSSGSSDSCSS  
SOSGSDSDSGSDSGSDSESDTSRENNVQAKPVPVDAEFKMSPSSTLAQRSML  
RKPOQAOOORPASNSGSESDSSGAKRKHNDEDMQSGSDSPQ  
SDSESEERDSSDCTESDYPKRVRSRCKNSKNGKKIIGOKKROIDSEDE  
DDEDYDNRKSSRRQATVAVSYKDEDEKTDSDDLLEYGCEVPQPEDEFETIERYM  
DCRVGRKATGATTTTVAVEDGDPAGFERKKEPEDIDYLLIKMGWSHINHWETEE  
TLQOQNVGRKRLDNYKRRKQDETRKMLKNASPEDVEYVQCDELDDLRKQYIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSWEDALISKFCOTCIDEYFSRNQSTPFK  
DCKVLKQRPFRVALKQPSYIGGHELELRDQINGLNLMLASHCMCKGNSCLADEMGL  
GKTQIOTISFLNLTFFHQLYGPPLVYPISTLSNOREIQMASOMNAVYVYGDINSR  
NMRTHEWHPQTKRLKFNILTTTEILIKDRAFLGGLMWATGDEARLKNDSLSL  
YKTLIDFKSNHRLITGTPLONSLKEMLSLHFIMPEKRSWEDEEHEGKRGEGYA  
SLHKELEPFLRRVYKDYKSLPAKVEQILRMEMSLQKQYKWLITRNYSKALSQSGK  
GSGSGFLNIMELKCKNCYILIKPDNNEFYNKQALQHLIRSSGKLTLLDKLIRL  
BERGNRVILFESQMWKLDITLAIYKRYRQPPORLDGSLKGLRKQALHFNAGSEDF  
CFLSTRAGGLGINTASADTYVTFPSDMNPQNDLQAKRAHRIQKQKQYNYRIVYTKG  
SVEEDILERRAKRMVLDHVIQRMQDTGTVLHTGSAPSSSTPNNKEELISALTKGAE  
ELFKEPEGEQEPQEMDIDELIKRAETHNEBGPLSVGDELLSQEVANFSNMDDDI



[illegible]

BASE COUNT	2130	a	1004	c	1243	g	1570	t	
ORIGIN									
Query Match	15.0%; Score 23; DB 9; Length 5947;								
Best Local Similarity	100.0%; Pred. No. 0.024;								
Matches	23;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	76	CAGACCGCTCGAGACTACCTCAT 98							
Db	4070	CAGACCGCTCGAGACTACCTCAT 4092							
RESULT 12									
AC092372/c									
LOCUS	AC092372	101220	bp	DNA	linear	PR1 07-DEC-2001			
DEFINITION	Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.								
ACCESSION	AC092372								
VERSION	AC092372.3	GI:17402768							
KEYWORDS	HTS.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.								
TITLE	1 (bases 1 to 101220)								
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.								
REFERENCE	Direct Submission								
AUTHORS	Unpublished								
TITLE	2 (bases 1 to 101220)								
JOURNAL	DOE Joint Genome Institute.								
REFERENCE	Direct Submission								
AUTHORS	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint								
TITLE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA								
JOURNAL	3 (bases 1 to 101220)								
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.								
AUTHORS	Direct Submission								
TITLE	Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell								
JOURNAL	Drive, Walnut Creek, CA 94598, USA								
COMMENT	On Dec 7, 2001 this sequence version replaced g1:15290448.								
	Draft Sequence Produced by DOE Joint Genome Institute								
	www.jgi.doe.gov								
	Finishing Completed at Stanford Human Genome Center								
	www.srhg.stanford.edu								
	Quality: Phrap Quality >=40 100% of Sequence;								
	Estimated Total Number of Errors is 0.								
	NOTE: This insert is not the entire sequence of the clone (entire								
	sequence is 146,740). It is clipped at the overlap with AC012624.								
	The number of bases overlapped is 90404.								
FEATURES									
source	1..101220								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/chromosome="5"								
	/clone="RP11-58M12"								
BASE COUNT	34122	a	18862	c	17827	g	30409	t	

```

ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 101220;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 76 CAGACCCGTGACAGACTACTCAT 98
Db 26887 CAGACCCGTGACAGACTACTCAT 26865

RESULT 13
AC012624 134365 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
DEFINITION AC012624
ACCESSION AC012624.6 GI:14993679
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
1. 134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTD-2082117"
/chromosome="5"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 134365;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 76 CAGACCCGTGACAGACTACTCAT 98
Db 118292 CAGACCCGTGACAGACTACTCAT 118314

RESULT 14
AC021449 143079 bp DNA linear HTG 10-SEP-2000
LOCUS Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces..
AC021449 143079 bp DNA linear HTG 10-SEP-2000
ACCESSION AC021449.3 GI:10047806
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 143079)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-58M12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143079)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Deavelano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Maddonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,D.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,K.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:1407963.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L5154
Center clone name: 58_M_12
-----
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134743 bases at least Q40
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 14400; agarose-fp
Insert size: 142179; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp
* 40512 43279: contig of 2768 bp in length
* 43280 43379: gap of 100 bp
* 43380 46905: contig of 3526 bp in length
* 46906 47005: gap of 100 bp
* 47006 51830: contig of 4825 bp in length
* 51831 51930: gap of 100 bp
* 51931 62619: contig of 10689 bp in length
* 62620 62719: gap of 100 bp
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp
* 75509 92516: contig of 17008 bp in length

```







PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 DR P-PSDB; AAM08147.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 XX birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 CC  
 CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758),  
 CC and the great tit CHD-W gene (see also AAT42759). Translated amino  
 CC acid sequences of this region are provided in AAM08146-49. The  
 CC CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 100.0%; Score 153; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-70;  
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTACTGATGATCGACAGAGAAAGCCAGCAAGCAGCTACAGACCAAGAAACC 60  
 DB 1 attttaccgtagatccagacaagaaccacccagcagaagcagctacagaccagaacc 60  
 OY 61 CAGGCAAGCAGCTACAGACCCGTCGAGACTACCTTAATTAATTAAGACTT 120  
 DB 61 cagcacaagcagctacagaccgctgcagactactcattactgaataagacctt 120  
 OY 121 GCAAGAAAGAGACCAAGAGCTTGTGCTGCA 153  
 DB 121 gcaagaagaagacacaaagcttctgtgtgca 153

RESULT 2  
 AAT42751  
 ID AAT42751 standard; cDNA; 6608 BP.  
 XX  
 AC AAT42751;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A gene.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 228..5390  
 FT /\*tag= a  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PE 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny  
 XX  
 XX Claim 1; Fig 5; 76pp; English.  
 PS  
 CC The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
 CC library using a great tit CHD-W sequence (see also AAT42755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AAT42756-57). It is located on an autosome or X  
 CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
 CC specific signal on hybridisation to genomic DNA of a non-ratite  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
 CC acids can also be used to control the sex of the progeny of a bird.  
 XX  
 SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 67.3%; Score 103; DB 18; Length 6608;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-44;  
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 CAAGAAACCCAGCAAGCAGCTACAGACCCGTCGAGACTACCTTAATTAATGAA 110  
 DB 4100 caagaacccccagcagaagcagctacagaccgctgcagactactcattactga 4159  
 OY 111 TAAAGACTTGCAGAAAGAGACACAAAGCTTGTGCTGCA 153  
 DB 4160 taagaccttgcagaagaagacacaaagcttctgtgtgca 4202

RESULT 3  
 AAT42758  
 ID AAT42758 standard; DNA; 153 BP.  
 XX  
 AC AAT42758;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-W gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52..81  
 FT /\*tag= a  
 FT /\*note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PE 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 XX  
 DR P-PSDB; AAM08148.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.

```

XX Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAAT42757),
CC chicken CHD-W (W refers to the W chromosome) gene (AAAT42758) and
CC and the great tit CHD-W gene (AAAT42759). Translated amino acid
CC sequences of this region are provided in AAW08146-49. The CHD-1A
CC (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX
SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match          30.7%; Score 47; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 7.1e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAAAGACGACA 137
   |||||||||||||||||||||||||||||||||||||||||||
DB 91 tactcctaataattactgataaagaaccttgcagaagaagacaca 137

RESULT 4
AAAT42754
ID AAAT42754 standard; CDNA; 1316 BP.
XX
AC AAAT42754;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chicken CHD-W gene (partial sequence).
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
OS Gallus sp.
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96MO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
PS WPI; 1997-043127/04..
XX
DR 1997-043127/04..
XX
FT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 1; Fig 8; 76pp; English.
XX
CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with
CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate
CC female development in birds. The sequence of CHD-W was deduced
CC from 2 clones isolated from a 10-day chick embryo library using
CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is
CC located on the W chromosome. Probes based on CHD-W and CHD-1A give
CC a W chromosome-specific signal on hybridisation to genomic DNA of a
CC non-rare bird and can be used for sex determin. of a bird. CHD-W
CC nucleic acids can also be used to control the sex of progeny of a
CC bird.
XX
SQ Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

Query Match          30.7%; Score 47; DB 18; Length 1316;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAAAGACGACA 137
   |||||||||||||||||||||||||||||||||||||||||||
DB 1198 tactcctaataattactgataaagaaccttgcagaagaagacaca 1244

RESULT 5
AAAT42759
ID AAAT42759 standard; DNA; 153 BP.
XX
AC AAAT42759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96MO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI; 1997-043127/04.
XX
DR P-PSDB; AAW08149.
XX
PT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAAT42757),
CC chicken CHD-W (W refers to the W chromosome) gene (AAAT42758) and
CC and the great tit CHD-W gene (AAAT42759). Translated amino acid
CC sequences of this region are provided in AAW08146-49. The CHD-1A
CC (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX
SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

Query Match          24.8%; Score 38; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 3.4e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAA 128
   |||||||||||||||||||||||||||||||||||||||
DB 91 tactcctaataattactgataaagaaccttgcagaaga 128

RESULT 6
AAAT42756
ID AAAT42756 standard; DNA; 153 BP.

```

AC	AAAT42756;
XX	
DT	12-MAR-1997 (first entry)
XX	
DE	Mouse CHD-1 gene (bases 3855-977).
KM	Bird; sex determination; chromodomain-Helicase-DNA binding <sup>1</sup> ;
KW	CHD-1; CHD-W; W chromosome; ss.
XX	
OS	Mus sp.
XX	
FH	Key Location/Qualifiers
FT	misc_difference 52..81
FT	/tag= a
FT	/note= "bases 52-81 are a repeat of bases 22-51
FT	and are ignored in the translated amino
FT	acid sequence given in Fig 3"
XX	
PN	WO9639505-A1.
XX	
PD	12-DEC-1996.
XX	
PE	05-JUN-1996; 96MO-CB01341.
XX	
PR	06-JUN-1995; 95GB-0011439.
XX	
PA	(ISIS-) ISIS INNOVATION LTD.
PI	Griffiths R, Tiwari B;
XX	
DR	WPI; 1997-043127/04.
DR	P-PSDB; AAM08146.
XX	
PT	Avian chromodomain-helicase-DNA binding genes determine sex in
PT	birds - used for sex determ. and to control sex of progeny
XX	
ES	Claim 8; Fig 3; 76pp: English.
XX	
CC	Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC	to portions of the chicken CHD-1A (A = Avian) gene (see also
CC	and AAT42757), chicken CHD-W (W refers to the W chromosome) gene
CC	(see also AAT42758) and the great tit CHD-W gene (see also AAT42759).
CC	Translated amino acid sequences of this region are provided in
CC	AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also
CC	AAT42754-55) genes determine sex in birds and can be used to identify
CC	the sex of an embryo, foetus etc. and to manipulate the sex of
CC	progeny.
XX	
SQ	Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;
XX	
Query Match	16.3%; Score 25; DB 18; Length 153;
Best Local Similarity	100.0%; Pred. No. 0.002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	74 TACAGACCCTGTCAGACTACCGCAT 98
Db	
	74 tacagaccctgcagactaccatccat 98
RESULT	7
AAVS9280	ID AAVS9280 standard; cDNA; 1311 BP.
XX	AAVS9280;
AC	
XX	14-DEC-1998 (first entry)
DT	
XX	
DE	Altered telomere repeat binding factor 1 gene.
KM	ds; human; telomere repeat binding factor; A-TFF; dimerisation domain;
KW	telomere; ageing; ataxia telangeictasia; Down's syndrome; tumour; viral.
XX	

OS	Homo sapiens.
OS	Synthetic.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1311
FT	/tag= a
FT	/product= "A-TRF"
XX	
FN	W09836066-A1.
PD	20-AUG-1998.
XX	
PE	13-FEB-1998; 98WO-US02765.
XX	
PR	04-FEB-1998; 98US-0018628.
PR	13-FEB-1997; 97US-0800264.
PA	(UYRO ) UNIV ROCKEFELLER.
PI	Bianchi A, De Lange T, Van Steensel B;
DR	WPI; 1998-480769/41.
DR	P-PSDB; AAM59280.
XX	
PT	Nucleic acid encoding altered telomere repeat binding protein and related vectors - transformants, hetero-dimers and antibodies, used to inhibit shortening of telomerases caused by ageing or disease,
PT	also used to extend life of cells in culture
XX	
PS	Claim 14; Page 110-111; 163pp; English.
XX	
CC	The altered vertebrate telomere repeat binding protein (A-TRF) has a
CC	telomere repeat binding factor (TRF) dimerisation domain, and forms a
CC	hetero-dimer with TRF, preventing it from binding to the specified repeat
CC	sequence. A-TRF, optionally expressed by gene therapy, is used to
CC	inhibit shortening of telomeres associated with ageing (for cosmetic
CC	purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,
CC	atrophy of the skin, age-related macular degeneration, atherosclerosis,
CC	tumours and viral (including human immune deficiency virus) infection.
CC	Cells expressing A-TRF also have an increased life span in vitro, e.g.
CC	for expression of recombinant proteins or where intended for subsequent
CC	transplant or for testing, eliminating the need for transformation.
CC	
SQ	Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other:
XX	
Query Match	16.3%; Score 25; DB 19; Length 1311;
Best Local Similarity	100.0%; Prod. No. 0.0021;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	74 TACAGACCGGTGCAGACTACTCAT 98
Db	1280 tacagaccctgacagactcattcat 1304
XX	
RESULT	8
ID	AAI80568/c
XX	AAI80568 standard; cDNA; 401 BP.
AC	AAI80568;
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide seq ID NO 628.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorders; arthritis; inflammation; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200164835-A2.



```

XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEO INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AA000637.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 628; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX cytokine cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, hematopoietic regulatory
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 401 BP; 113 A; 87 C; 108 G; 93 T; 0 other;
XX
XX Query Match 11.8%; Score 18; DB 22; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 9;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 17 CAGACAGAAACCCGAG 34
XX ||||||||||||||||
XX 304 CAGACAGAAACCCGAG 287
XX
XX RESULT 9
XX ID AAF68219 standard; cDNA; 552 BP.
XX AC AAF68219;
XX
XX 12-APR-2001 (first entry)
XX
XX Human lung tumour protein related nucleotide sequence SEQ ID NO:137.
XX
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX cytostatic; antisense inhibition; ss.
XX
XX Homo sapiens.
XX OS
XX WO200100828-A2.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000MO-US18061.
XX
XX 30-JUN-1999; 99US-0346492.
XX 15-OCT-1999; 99US-0419356.
XX 17-DEC-1999; 99US-0466867.

```

```

PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-2000; 2000US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546259.
PR 27-APR-2000; 2000US-0560406.
PR 05-JUN-2000; 2000US-0589184.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX Retter MW, Mannion J;
XX WPI; 2001-071488/08.
XX
XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX useful for preventing, diagnosing and treating lung cancer -
XX
XX Example 1; Page 193; 436pp; English.
XX
XX The present invention describes immunogenic portions of lung tumour-
XX associated proteins (I) and the nucleic acids (NAs) that encode them.
XX (I) have cytostatic activity and can be used in gene therapy, antisense
XX inhibition and in vaccines. The NAs and the lung tumour-associated
XX proteins they encode may be used in the prevention, treatment and
XX diagnosis of diseases associated with their inappropriate expression,
XX especially lung cancers. For example, the NAs may be administered to
XX treat diseases by rectifying mutations or deletions in a patient's genome
XX that affect the activity of the protein by expressing inactive proteins
XX or to supplement the activity of the protein production of (I). Additionally, the
XX NAs may be used to produce the lung-tumour associated protein, according
XX to standard recombinant DNA methodology. Conversely, antisense NA
XX molecules may be administered to down regulate protein expression by
XX binding with the cells own genes and preventing their expression. The NA
XX and complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar NA sequences in
XX cancer. The (I) may be used as antigens in the production of antibodies
XX and in assays to identify modulators (agonists and antagonists) of the
XX expression and activity of the protein. AAF68083 to AAF68878 and
XX AAF68848 to AAF68878 represent human lung tumour protein related
XX nucleotide and protein sequences which are used in the exemplification
XX of the present invention.
XX
XX Sequence 552 BP; 220 A; 118 C; 69 G; 144 T; 1 other;
XX
XX Query Match 11.8%; Score 18; DB 22; Length 552;
XX Best Local Similarity 100.0%; Pred. No. 9;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 97 ATTAATTCTGATATA 114
XX ||||||||||||||||
XX 492 attaatctactcgataaa 509
XX
XX RESULT 10
XX ID AAV49574 standard; cDNA to mRNA; 1409 BP.
XX AC AAV49574;
XX
XX 21-OCT-1998 (first entry)
XX
XX Human lymphoma cell line U937 clone HP10136 cDNA #1.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
XX differentiation; immune system; stimulator; suppressor; regulator;
XX hematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
XX haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
XX Homo sapiens.
XX OS
XX
XX Key Location/Qualifiers

```

```
FT      CDS          82..729  
FT      /*tag=       a  
XX      /product= "transmembrane domain containing protein"  
PN      MO982J328-A2.  
XX      22-MAY-1998.  
XX      PD  
XX      PF      07-NOV-1997;    97WO-JP04056.  
XX      PR      13-NOV-1996;    96JP-0301429.  
XX      PA      (PROT-) PROTEGENE INC.  
XX      PA      (SAGA ) SAGAMI CHEM RES CENTRE.  
XX      PI      Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
XX      DR      WPI: 1998-297932/26.  
XX      DR      P-PsDB: AAM64546.  
XX      PT      Human protein having transmembrane domain - useful for, e.g.  
XX      PT      research and nutrition  
XX      PS      Claim 4; Page 153-155; 205pp; English.  
XX      AAAY9550-Y49599 are cDNA sequences which encode human proteins containing  
CC      a transmembrane domain. These proteins can be used for, e.g. research  
CC      and nutrition, and may have cytokine and cell  
CC      proliferation/differentiation, immune stimulating/suppressing,  
CC      haematopoiesis regulating, tissue growth, activin/inhibn,  
CC      chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,  
XX      anti-inflammatory or tumour inhibition activity.  
XX      SQ      Sequence 1409 BP; 377 A; 278 C; 325 G; 429 T; 0 other:  
  
Query Match          11.8%; Score 18; DB 19; Length 1409;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  
  
QY      111 TTAAGACCTTGCAAGAA 128  
        |||||||  
DB      842 taaagaccttgcaagaaa 859  
  
RESULT 11  
ID      AAX56240  
XX      AC      AAX56240 standard; CDNA; 1462 BP.  
XX      AC      AAX56240;  
XX      DT      16-JUL-1999 (first entry)  
DE      Human CBFBBA01 vesicle trafficking protein SEC22b gene.  
XX      KW      Human; vesicle trafficking protein; SEC22b; CBFBBA01; cancer;  
KW      autoimmune disease; diabetes mellitus; multiple sclerosis;  
XX      diagnosis; ss.  
XX      OS      Homo sapiens.  
XX      OS      WO9921988-A1.  
XX      PN      PD      06-MAY-1999.  
XX      PF      29-OCT-1997;    97WO-CN00115.  
XX      PR      29-OCT-1997;    97WO-CN00115.  
XX      PA      (UYSH-) UNIV SHANGHAI SECOND MEDICAL.  
XX      PI      Wang Y, Wu J;
```

DR		WPI: 1999-312961/26.
DR	P-PSDB:	AAV09512.
XX	CBFBA01	protein useful in treating cancer, autoimmune disease,
PT		diabetes mellitus and multiple sclerosis
XX		
PS	Claim 2;	Page 21; 29pp; English.
XX		
CC	The present sequence encodes CBFBA01, which is a protein homologous to	
CC	murine vesicle trafficking protein sec22b. CBFBA01 is a	
CC	cytoplasmically-oriented integral membrane protein, located in the	
CC	endoplasmic reticulum and golgi membranes. CBFBA01 polynucleotide (I),	
CC	vectors containing (I) and recombinant host cells are useful for	
CC	recombinant production of CBFBA01. Both (I), CBFBA01 and antibodies	
CC	against CBFBA01 are useful as research reagents, for screening assays	
CC	and in diagnostic assays, especially for cancer, autoimmune disease,	
CC	diabetes mellitus and multiple sclerosis. Antagonists and agonists of	
CC	CBFBA01 can be used to inhibit or enhance, respectively, the activity	
CC	of CBFBA01 or expression of (I). Anti-CBFBA01 antibodies and CBFBA01	
CC	or its fragments can be used in vaccines.	
SO	Sequence 1462 BP: 419 A; 286 C; 323 G; 434 T; 0 other:	
OY	111 TTAACAGCTTGCACAAA 128	
Db	826 taaagaccttgcaagaaa 843	
RESULT 12		
AAV20467/C		
ID	AAV20467 standard; DNA: 3225 BP.	
XX		
AC	AAV20467:	
XX		
DT	17-JUN-1998 (first entry)	
DE	Human c-myb oncogene.	
XX		
KW	Human; oncogene; proto-oncogene; neoplastic disease; anticancer;	
OS	cancer; antisense oligonucleotide; c-myp; ds.	
XX	Homo sapiens.	
PN	US5734039-A.	
PD	31-MAR-1998.	
PF	15-SEP-1994; 94US-0306691.	
PR	15-SEP-1994; 94US-0306691.	
PA	(UYJE-) UNIV JEFFERSON THOMAS.	
PI	Calabretta B, Skorski T;	
DRI	WPI: 1998-229882/20.	
PT	Anticancer composition comprising two anti-sense oligo:nucleotide(s)	
XX	- Targetting cytoplasmic and nuclear oncogene(s)	
PS	Claim 1; Column 131-134; 92pp; English.	
CC	The present sequence represents an oncogene from the present invention.	
CC	The present invention describes a composition which comprises two	
CC	antisense oligonucleotides. The first oligonucleotide is specific for a	
CC	cytoplasmic oncogene or proto-oncogene selected from ras, raf, Egr-1,	
CC	c-fms, c-fos, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and	
CC	c-ves. The second oligonucleotide is specific for a nuclear oncogene or	

CC proto-oncogene selected from myc, jun, c-ets, c-fos, c-myc, B-myb,  
 CC c-rel, c-vav, c-ski, c-spl, cyclin D1, pml/rar alpha, AML1/MTG8,  
 CC E2A/P1 and ALL-1/AF-4. The composition is used for treating cancer.  
 CC The combination of antisense oligonucleotides has synergistically  
 CC enhanced ability to inhibit growth of cancer cells.  
 XX  
 SO Sequence 3225 BP; 975 A; 687 C; 698 G; 865 T; 0 other;

Query Match 11.8%; Score 18; DB 19; Length 3225;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97.ATTAAATTACTGAAATRAA 114  
 |||  
 DB 2662 ATTAAATTACTGAAATRAA 2645

RESULT 13  
 ABL15492/C  
 ID ABL15492 standard; cDNA: 59967 BP.  
 XX  
 AC ABL15492;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide seq ID NO 40958.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 OS  
 XX Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PE 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PA  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB71389.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 40958; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL20511), expressed DNA  
 CC sequences (ABB57737-ABB72072).  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SO Sequence 59967 BP; 17305 A; 12463 C; 12291 G; 17908 T; 0 other;

Query Match 11.8%; Score 18; DB 23; Length 59967;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ACAGAAACCCAGACCA 37  
 |||  
 DB 21570 ACAGAAACCCAGACCA 21553

RESULT 14  
 AAZ01319  
 ID AAZ01319 standard; DNA: 19 BP.  
 XX  
 AC AAZ01319;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX  
 DE PCR primer for PGI biallelic marker 99-140-130.  
 XX  
 KM PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
 KM cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
 KM PSA; human; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 PN WO932644-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PE 22-DEC-1998; 98WO-IB02133.  
 XX  
 PR 09-SEP-1998; 98US-0099658.  
 PR 22-DEC-1997; 97US-0996306.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
 XX  
 DR WPI; 1999-405178/34.  
 XX  
 PT Use of a prostate cancer associated gene and biallelic markers  
 PT derived from it  
 XX  
 PS Claim 4; Page 368; 385pp; English.  
 XX  
 CC The invention relates to a mammalian PGI gene and protein, and a set of  
 CC PGI biallelic markers. The PGI polynucleotide and biallelic markers are  
 CC used in a hybridisation assay, a sequencing assay, or in an  
 CC allele-specific amplification assay for determining the identity of a  
 CC nucleotide at a PGI-related biallelic marker. The methods can be used to  
 CC detect and to assess the risk of developing cancer or prostate cancer.  
 CC Early-stage diagnosis of prostate cancer relies on prostate specific  
 CC antigen (PSA) dosage. However, the effectiveness of this is limited due  
 CC to its inability to discriminate between malignant and non-malignant  
 CC affections of the organ. A need exists for both a reliable diagnostic  
 CC procedure which would enable early-stage diagnosis, and for preventative  
 CC and curative treatments of the disease. The PGI gene can be used for  
 CC detection of prostate cancer, and the risk of developing it in the  
 CC future, and can also be used to determine therapies for the disease.  
 XX  
 SO Sequence 19 BP; 9 A; 6 C; 3 G; 1 T; 0 other;

Query Match 11.8%; Score 17; DB 20; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAAGCAGCTACGACCA 52  
 |||  
 DB 3 aaagcagctacgaccca 19

RESULT 15  
 AAZ01079  
 ID AAZ01079 standard; DNA: 47 BP.

XX AA201079;  
AC XX  
XX XX  
DT 27-SEP-1999 (first entry)  
XX XX  
DE Probe for human PGI biallelic marker 99-140-130.  
XX XX  
XX PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
KW PSA; human; ss.  
XX XX  
OS Synthetic.  
OS Homo sapiens.  
XX XX  
PN WO932644-A2.  
XX XX  
PD 01-JUL-1999.  
XX XX  
PF 22-DEC-1998; 98WO-IB02133.  
XX XX  
PR 09-SEP-1998; 98US-0099658.  
PR 22-DEC-1997; 97US-0996306.  
XX XX  
PA (GEST ) GENSET.  
XX XX  
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
XX XX  
DR WPI; 1999-405178/34.  
XX XX  
PT Use of a prostate cancer associated gene and biallelic markers  
PI derived from It  
XX XX  
PS Claim 4; Page 313; 385pp; English.  
XX XX  
SQ

The invention relates to a mammalian PGI gene and protein, and a set of PGI biallelic markers. The PGI polynucleotide and biallelic markers are used in a hybridisation assay, a sequencing assay, or in an allele-specific amplification assay for determining the identity of a nucleotide at a PGI-related biallelic marker. The methods can be used to detect and to assess the risk of developing cancer or prostate cancer. CC Early-stage diagnosis of prostate cancer relies on prostate specific antigen (PSA) dosage. However, the effectiveness of this is limited due to its inability to discriminate between malignant and non-malignant affections of the organ. A need exists for both a reliable diagnostic procedure which would enable early-stage diagnosis, and for preventative and curative treatments of the disease. The PGI gene can be used for detection of prostate cancer, and the risk of developing it in the future, and can also be used to determine therapies for the disease.

Query Match 11.1%; Score 17; DB 20; Length 47;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAAGCAGCTACAGACCA 52  
|||||  
DB 7 aaagcagctacagacca 23

Search completed: August 3, 2002, 01:50:07  
Job time: 32192 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:33 ; Search time 7016.61 Seconds

(without alignments)  
294.307 Million cell updates/sec

Title: US-08-973-363-3

Sequence: 1 ATTTCACCTGATGATCCAGA.....CACAAAGCCTGCTGCTGCA 153

Scoring table:

ORIGO\_NUC

Gapop 60.0 , Gapext 60.0

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em.estba:\*  
2: em.esthum:\*  
3: em.estlu:\*  
4: em.estmu:\*  
5: em.estov:\*  
6: em.estpl:\*  
7: em.estro:\*  
8: em.estl:\*  
9: gb.estl:\*  
10: gb.est2:\*  
11: gb.hic:\*  
12: gb.gss:\*  
13: em.gss.hum:\*  
14: em.gss.liv:\*  
15: em.gss.pln:\*  
16: em.gss.vtl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	16.3	619	9	BB155356
2	25	16.3	619	9	BB155356
3	23	15.0	430	9	BB461065
4	23	15.0	430	9	AL601246
5	23	15.0	438	9	BB830730
6	23	15.0	446	9	BB834922
7	23	15.0	547	9	AI890775
8	23	15.0	686	9	AI890758
9	23	15.0	821	10	BF239667
10	23	15.0	866	9	BF239667
11	21	13.7	1028	10	BE895133
12	20	13.1	664	12	A2570983
13	20	13.1	693	12	BB351821
14	19	12.4	505	12	AG128961
15	19	12.4	527	12	AQ453778
16	19	12.4	558	12	A2486207
17	19	12.4	573	12	A2392020
					AQ988250
					RPCT-23-3

18	19	12.4	695	12	A2963945
19	19	12.4	895	10	BE971206
20	19	12.4	944	12	A2200903
21	18	11.8	110	10	221483
22	18	11.8	157	12	BB4318
23	18	11.8	167	10	BE842996
24	18	11.8	225	9	AV317344
25	18	11.8	225	9	AV317684
26	18	11.8	233	9	AV312872
27	18	11.8	233	9	AV375787
28	18	11.8	233	9	BB164147
29	18	11.8	276	9	AV222464
30	18	11.8	276	10	BE768484
31	18	11.8	291	9	BB030255
32	18	11.8	293	9	BB031183
33	18	11.8	301	9	BB402744
34	18	11.8	302	9	BB560958
35	18	11.8	303	9	BB17675
36	18	11.8	314	9	AA354813
37	18	11.8	322	9	AA957817
38	18	11.8	322	10	BF289372
39	18	11.8	323	9	BB137014
40	18	11.8	355	9	AI902726
41	18	11.8	416	9	BB810785
42	18	11.8	424	12	AO816352
43	18	11.8	426	9	AA267899
44	18	11.8	437	9	AA955076
45	18	11.8	438	9	AA234796

## ALIGNMENTS

RESULT 1  
BB155356  
LOCUS  
DEFINITION  
BB155356 RIKEN full-length enriched, 16 days neonate thymus Mus  
musculus cDNA clone A130024L16.3' similar to I10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION  
BB155356  
VERSION  
BB155356.2  
GI:16268254

KEYWORDS  
house mouse,  
mus musculus

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

Akakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,  
'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
'D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T.,  
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Akakawa,T., et al. 2001)  
Unpublished (2001)

On Jun 29, 2000 this sequence version replaced gi:881186.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
'M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura



		GAGGAGAGATTCCTGCAGTTAAATTAAATTAATGACCCCGCCCCC 3'}. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambda F1C I."	
BASE COUNT	211 a 130 c 168 g 151 t		
ORIGIN			
Query Match	16 3%: Score 25; DB 9; Length 660;		
Best Local Similarity	100.0%: Pred. No. 0.0062;		
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Oy	74 TACGACCCCTGCAGACTACTCAT 98		
Db	524 TACGACCCCTGCAGACTACTCAT 548		
RESULT 3			
LOCUS	AL601246 430 bp mRNA linear EST 14-AUG-2001		
DEFINITION	DKFZP313J1040_r1 313 (synonym: h1cc2) Homo sapiens cDNA clone		
ACCESSION	AL601246		
VERSION	AL601246.1 GI:15164752		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 430)		
TITLE	Bloeker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann		
JOURNAL	5, Est (Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Well, B. and Wiemann, S.)		
COMMENT	Unpublished (1999)		
	Contact: Bloecker H		
	MPs		
	Am Klopferpitz 1a, D-82152 Martinsried, Germany		
	This is the 5' sequence of the clone insert		
	clone from S. Wiemann, Molecular Genome Analysis, German Cancer		
	Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; Ltd,		
	sequenced by GBR (National Research Centre for Biotechnology of the		
	Braunschweig/Germany) within the cDNA sequencing consortium of the		
	German Genome Project.		
	No st sequence available.		
	This clone (DKFZP313J1040) is available at the RZPD in Berlin.		
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059		
	Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.		
FEATURES	Location/Qualifiers		
source	1..430		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKFZP313J1040"		
	/clone_1lb="313 (synonym: h1cc2)"		
	/dev_stage="adult"		
	/lab_host="DH10B"		
	/note="Vector: pT7p1ex2; site_1: sf11a; site_2: sf11b;		
	cDNA-collection"		
BASE COUNT	163 a 81 c 86 g 100 t		
ORIGIN			
Query Match	15 0%: Score 23; DB 9; Length 430;		
Best Local Similarity	100.0%: Pred. No. 0.071;		
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Oy	76 CAGACCCCTGCAGACTACTCAT 98		
Db	263 CAGACCCCTGCAGACTACTCAT 285		
RESULT 4			
LOCUS	BB830730 438 bp mRNA linear EST 19-NOV-2001		

DEFINITION	BB830730 RIKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA Mus musculus CDNA Clone G930013K04 3', mRNA			
ACCESSION	BB830730			
VERSION	BB830730.1			
KEYWORDS	EST			
SOURCE	house mouse			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 438)			
AUTHORS	Akimura,T., Hirawo,T., Carninci,P., Furuno,M., Hanagaki,T., Akimura,T., Arakawa,K., Hirooka,T., Hirozane,T., Imclari,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shimagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,T., Takaku-Akahita,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)			
TITLE	Unpublished (2001)			
JOURNAL	Contact: Yoshinide Hayashizaki			
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 220-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, url:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayashi,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Watahiki,K., Fujiyake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,K., Yoneda,T., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,T., Kawai,J., Okazaki,T., Muramatsu,M., Inoue,T., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details. e mouse tissues.			
FEATURES	Location/Qualifiers			
source	1..438			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="G930013K04"			
	/clone_1b="YRKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA"			
	/tissue_type="mammary gland"			
	/cell_line="RCB-0527 Jyg-MC(B)"			
BASE COUNT	138 a 96 c 108 g 96 t			
ORIGIN				
Query Match	15.0%; Score 23; DB 9; Length 438;			
Best Local Similarity	100.0%; Pred. No. 0.071;			
Matches	23; Conservative 0; Mismatches 0; Indels 0; Caps 0;			
74	TACGACCGCTGCACACTCTC 96			
343	TACGACCGCTGCACACTCTC 365			

RESULT 5  
BB834922 446 bp mRNA linear EST 19-NOV-2001  
LOCUS BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
DEFINITION Jy9-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.  
ACCESSION BB834922  
VERSION BB834922.1 GI:17013165  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
AUTHORS Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Matshiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.riken.go.jp/  
URL: http://genome-gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagl,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome-gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

FEATURES  
Location/Qualifiers  
source 1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033J21"  
/clone.lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"  
BASE COUNT 139 a 99 c 108 g 100 t  
ORIGIN  
Query Match 15.0%; Score 23; DB 9; Length 446;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

74 TACAGACCGTGACACTACCTC 96  
|||||

Db 352 TACAGACCGTGACACTACCTC 374

RESULT 6  
A1890775/c 547 bp mRNA linear EST 07-MAR-2000  
LOCUS A1890775 NCI-CGAP ut2 Homo sapiens cDNA clone IMAGE:2443725 3',  
DEFINITION similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapdb-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILIN at:  
[www.bio.lnl.gov/bdrp/image/image.html](http://www.bio.lnl.gov/bdrp/image/image.html)  
Insert Length: 1924 Std Error: 0.00  
Seq primer: -40up from G1bco  
High quality sequence stop: 418.

FEATURES  
Location/Qualifiers  
source 1..547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone.lib="NCI-CGAP ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; site\_1: SalI;  
site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 114 a 118 c 85 g 230 t  
ORIGIN  
Query Match 15.0%; Score 23; DB 9; Length 547;  
Best Local Similarity 100.0%; Pred. No. 0.074;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

76 CAGACCCGTGCAGACTACCTCAT 98  
|||||

Db 333 CAGACCCGTGCAGACTACCTCAT 311

RESULT 7  
AW997058/c 686 bp mRNA linear EST 05-JUN-2000  
LOCUS AW997058 OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW997058  
VERSION AW997058.1 GI:8257292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,



Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**TITLE**  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
**JOURNAL**  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
**MEDLINE**  
 20202663  
**COMMENT**  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPBS/PCR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=kt2-QV3-BN0047-150  
 400-152-c03&f3=2000-04-15&f4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site: 1: SmaI;  
 Site: 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN

Query Match 15.0%; Score 23; DB 9; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 0.077; 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;

Oy 76 CAGACCCGTCGACGACTCCTCAT 98  
 ||||||||||||||||||||  
 Db 83 CAGACCCGTCGACGACTCCTCAT 61

**RESULT** 8  
 BF239967 821 bp mRNA linear EST 14-NOV-2000  
 LOCUS 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@strs@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

Found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LICM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1..821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="IMAGE:4133129"  
 /clone\_lib="NIH\_MGC\_54"  
 /tissue\_type="from Chronic myelogenous leukemia"  
 /lab\_host="pMD108 (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pMDR-LIB (Clontech);  
 Site: 1: SfiI (ggccgctggcgc); Site: 2: SfiI (ggccataggc  
 ); Double-stranded cDNA was prepared from cell line RM.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-ATTCTAGAGCGCCAGCGCCGACATG-dt(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN

Query Match 15.0%; Score 23; DB 10; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 0.079; 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;

Oy 76 CAGACCCGTCGACGACTCCTCAT 98  
 ||||||||||||||||||||  
 Db 33 CAGACCCGTCGACGACTCCTCAT 55

**RESULT** 9  
 AU125712 866 bp mRNA linear EST 23-OCT-2000  
 LOCUS AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
 DEFINITION Sequence.  
 ACCESSION AU125712  
 VERSION AU125712.1 GI:10950428  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
 Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.  
 HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
 Y., Sugano,S., Isogai,T.)  
 Unpublished (2000)  
 Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp  
 HRI human cDNA project: 5'- 6 3'-end one pass sequencing: Helix  
 Research Institute, cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers  
 1..866  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RM4002061"  
 /clone\_lib="NT2RM4"

```

/cell_type="teratocarcinoma"
/cell_line="NT2"
/Note="Vector: PME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"
BASE COUNT      312 a      149 c      196 g      207 t      2 others
ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 866;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCGAGACTACTCAT 98
|||||
Db 495 CAGACCCGTCGAGACTACTCAT 517

RESULT 10
BE895133      1028 bp      mRNA      linear      EST 20-OCT-2000
LOCUS      601436060F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3921087 5',
DEFINITION      mRNA sequence.
ACCESSION      BE895133
VERSION      BE895133.1 GI:10358221
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 1028)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cga@bts-remail.nih.gov
Tissue Procurement: ATCC/DCID/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM9753 row: h column: 16
High quality sequence stop: 488.
Location/Qualifiers
1. 1028
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921087"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: PCMV-SpO6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
technologies."
BASE COUNT      387 a      205 c      238 g      198 t
ORIGIN

Query Match      15.0%; Score 23; DB 10; Length 1028;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCGAGACTACTCAT 98
|||||
Db 176 CAGACCCGTCGAGACTACTCAT 198

RESULT 11
A2570983      664 bp      DNA      linear      GSS 15-MAY-2001
LOCUS      281PVC04 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
DEFINITION

```

```

ACCESSION      A2570983
VERSION      A2570983.1 GI:13982617
KEYWORDS      GSS.
SOURCE      malaria parasite P. vivax.
ORGANISM      Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE      1 (bases 1 to 664)
AUTHORS      Carlton J.M.-R. and Dame J.B.
TITLE      The plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL      Parasitol. Today 16 (10), 409 (2000)
COMMENT      Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.
Location/Qualifiers
1. 664
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
497-598)"
/db_xref="taxon:5855"
/clone_lib="PV MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/Note="Vector: pBluescript SK(+) vector DNA, phagemid
excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidpur
filter, followed by passage through a column of pre-wet
Whatman CF11 powder (1:2 ratio volume of blood to CF11),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nuclease
in the presence of 44% formamide at 500c as described
(Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA Polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
500bp-4kb were ligated into the Eco RV site of plusscript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."
BASE COUNT      197 a      159 c      115 g      193 t
ORIGIN

Query Match      13.7%; Score 21; DB 12; Length 664;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GAAAGGAGGACCAAGGCTTG 145
|||||
Db 402 GAAAGGAGGACCAAGGCTTG 422

RESULT 12
BH351821      693 bp      DNA      linear      GSS 03-DEC-2001
LOCUS      CH230-160016.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
DEFINITION      CH230-160016, DNA sequence.
ACCESSION      BH351821
VERSION      BH351821.1 GI:17282555
KEYWORDS      GSS.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 693)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 EcorI segment
JOURNAL Unpublished (1999)
COMMENT Other-GSSs: CH230-160016.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources (pdejong@mail.cho.org).
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 160 Row: 0 Column: 16
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
1..693
/organism="Rattus norvegicus"
/strain="BN/SSNhsd/McW"
/db_xref="taxon:10116"
/clone="CH230-160016"
/clone_1lb="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: PTARBAC2.1; Site_1: EcorI; Site_2: EcorI;
CHORI-230 Rat (BN/SSNhsd/McW) BAC library produced by
Pieter de Jong"
BASE COUNT 285 a 147 c 158 g 103 t
ORIGIN
Query Match 13.1%; Score 20; DB 12; Length 693;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 122 CAGGAGGAGGACGACGAGG 141
Db 597 CAGGAGGAGGACGACGAGG 616
RESULT 13
AG128961 838 bp DNA linear GSS 04-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-140E13.F, genomic survey sequence.
ACCESSION AG128961
VERSION AG128961.1 GI:16658126
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
BAC library clone:PTB-140E13.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
1 (sites)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 838)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpesegsc.riken.go.jp, url:http://hgsc.gsc.riken.go.jp/,

```

```

COMMENT Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pRS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..838
/organism="Pan troglodytes"
/db_xref="taxon:9538"
/clone="PTB-140E13.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1lb="PTB Chimpanzee Male BAC library"
BASE COUNT 286 a 276 c 177 g 44 t 55 others
ORIGIN
Query Match 13.1%; Score 20; DB 12; Length 838;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 24 GAACCCGAGGCAAGCAGC 43
Db 458 GAACCCGAGGCAAGCAGC 477
RESULT 14
AQ453778 505 bp DNA linear GSS 21-APR-1999
LOCUS HS_5170.AL.D02.T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=746 Col=3 Row=G, DNA sequence.
ACCESSION AQ453778
VERSION AQ453778.1 GI:4594948
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 505)
Mahitras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,D., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahitras GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pdejong@med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 746 Row: G Column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 505.
Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=746 Col=3 Row=G"

```

```

/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/Note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT      197 a      77 c      107 g      119 t      5 others
ORIGIN

Query Match      12.4%; Score 19; DB 12; Length 505;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 AAGAAGGAGCACAAGG 141
Db 132 AAGAAGGAGCACAAGG 150

RESULT 15
AZ486207/c 527 bp DNA linear GSS 05-OCT-2000
LOCUS 1M0314M02F Mouse 10kb plasmid UGCC1M library Mus musculus genomic
DEFINITION
ACCESSION AZ486207
VERSION AZ486207.1 GI:10652755
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Mus;
1 (bases 1 to 527)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Petersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10Kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., StC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0314 Row: M Column: 02
Seq primer: CGTGTGTAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 527.
Location/Qualifiers
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCC1M0314M02"
/clone_lib="Mouse 10Kb plasmid UGCC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/Note="Vector: PMD42uv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel

```

```

electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      174 a      88 c      87 g      178 t
ORIGIN

Query Match      12.4%; Score 19; DB 12; Length 527;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GAAGCACAAGGCTGCTG 148
Db 404 GAAGCACAAGGCTGCTG 386

```


Search completed: August 2, 2002, 22:41:37  
Job time: 30152 sec

Mon Aug 5 11:51:50 2002

us-08-973-363-3.011.rst

---

Page 9







Query Match 100.0%; Score 153; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.2e-66;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTTTACCTGATGACCCAGATAAGAAACACAGCAAGAGTTCGACACCAAGAAACCA 60  
|||||  
Db 1 ATTTTACCTGATGACCCAGATAAGAAACACAGCAAGAGTTCGACACCAAGAAACCA 60

Qy 61 CAGGCAAGAGAGTTGACAGACCCGTCGAGATTACTCTTAATTACTGATTAAGACCTT 120  
|||||  
Db 61 CAGGCAAGAGAGTTGACAGACCCGTCGAGATTACTCTTAATTACTGATTAAGACCTT 120

Qy 121 GCAGAAAGAGTGCAGAAAGCTTACTGTGCA 153  
|||||  
Db 121 GCAGAAAGAGTGCAGAAAGCTTACTGTGCA 153

RESULT 2  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN

Query Match 34.6%; Score 53; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 3.9e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 CAGACCGTGACAGATTACTCTTAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
Db 76 CAGACCGTGACAGATTACTCTTAATTACTGATTAAGACCTTGCAGAGAA 128

RESULT 3  
LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 15 from Patent WO9639505.  
ACCESSION A58696  
VERSION A58696.1 GI:3714253  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1316)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 15 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .1316  
/organism="unidentified"

BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN /db\_xref="taxon:32644"

Query Match 30.7%; Score 47; DB 6; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 CGTGCAGATTACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
Db 1189 CGTGCAGATTACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 1235

RESULT 4  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

Query Match 24.8%; Score 38; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 TACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
Db 91 TACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128

RESULT 5  
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999  
DEFINITION delta-crystallin enhancer binding protein, complete sequence.  
ACCESSION D14316  
VERSION D14316.1 GI:391639  
KEYWORDS  
SOURCE Gallus gallus (library: lambda gt11) 13 day embryo lens cDNA to mRNA, clone vfl1.  
ORGANISM Gallus gallus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
AUTHORS 1 (bases 1 to 2292)  
TITLE Funahashi,J.  
JOURNAL Direct Submission  
SUBMITTED (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel., Aging and Cancer, Tohoku Univ.; 4-1 Seiry-machi, Aoba-ku, Sendai 980-77,  
Japan (Tel:022-272-9499, Fax:022-272-3982)  
2 (sites)  
REFERENCE Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.  
AUTHORS Delta-crystallin enhancer binding protein delta EFL is a zinc  
TITLE finger-homeodomain protein implicated in postgastrulation embryogenesis  
JOURNAL Development 119 (2), 433-446 (1993)





```

misc-feature      4341..4604
                  /gene="CHD-2"
                  /note="short insert found in longer variant mRNA of CHD-2"
BASE COUNT      2446 a 1223 c 1520 g 1683 t
ORIGIN

Query Match      24.8%; Score 38; DB 5; Length 6872;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTTACTGATAAGACCTTGCAAGAA 128
      |||||||
Db 4140 TACCTCATTAATTTACTGATAAGACCTTGCAAGAA 4177

RESULT 8
LOCUS      MM091538      959 bp      mRNA      linear      ROD 12-APR-1997
DEFINITION Mus musculus vesicle trafficking protein sec22b mRNA, complete cds.
ACCESSION  U91538
VERSION     U91538.1 GI:1907385
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 959)
AUTHORS    Hay, J.C., Chao, D.S., Kuo, C.S. and Scheller, R.H.
TITLE      Protein interactions regulating vesicle transport between the
            endoplasmic reticulum and Golgi apparatus in mammalian cells
            Cell 89 (1), 149-158 (1997)
JOURNAL    97248495
MEDLINE    2 (bases 1 to 959)
REFERENCE   Hay, J.C., Chao, D.S., Kuo, C.S. and Scheller, R.H.
AUTHORS    Direct Submission
TITLE      Submitted (27-FEB-1997) Molecular and Cellular Physiology, Stanford
            University, B155 Beckman Center, Stanford, CA 94305, USA
FEATURES
    source
        1..959
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /tissue_type="Placenta"
            77..724
            /note="similar to Sec22"
            /codon_start=1
            /product="vesicle trafficking protein sec22b"
            /protein_id="AAC53130.1"
            /db_xref="gi:1907386"
            /translation="AVLLMTIARVADGLPLASMOEDSGRDLQOYOSQAKLPRL
            NEOSPRTCTAGAMTFHYITIOGCVYLCEAAPPKLAFAYLEDLSEPEQHGK
            VPTVSRPSFIETDFDIOTKTKLYIDSRARRLGSINTELQDVQIMVANIIEVYORG
            EALSRLDSKANLSTLSKKRYODAKYLMKRSYAKLAAYAVFIMLYIVRRWML"
CDS
    BASE COUNT      241 a 219 c 261 g 238 t
ORIGIN

Query Match      13.7%; Score 21; DB 10; Length 959;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 TAAAGACCTTGCAAGAAAGA 131
      |||||||
Db 835 TAAAGACCTTGCAAGAAAGA 855

RESULT 9
LOCUS      BC009024      1339 bp      mRNA      linear      ROD 12-JUL-2001
DEFINITION Mus musculus; SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1, clone MGC:5885 IMAGE:3481731, mRNA, complete
            cds.

```

```

ACCESSION      BC009024
VERSION        BC009024.1 GI:14290511
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1339)
AUTHORS    Strausberg, R.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK
COMMENT       NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgeabs-remail.nih.gov
            Tissue Procurement: Lotnar Hennighausen Ph.D., Robin Humphreys
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLJ)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalona@bcm.tmc.edu.
            Villalón, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
            Muzny, D.M., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LNLJ at: http://image.llnl.gov
            Series: IRAX Plate: 6 Row: K Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 1907385.
FEATURES
    source
        1..1339
            /organism="Mus musculus"
            /db_xref="LocusID:9554"
            /db_xref="taxon:10090"
            /clone="MGC:5885 IMAGE:3481731"
            /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
            old, gross tissue."
            /clone_11b="NCI-CGAP_Mam5"
            /lab_host="DH10B"
            /note="Vector: PCMV-SPORT6"
            82..729
            /codon_start=1
            /product="SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1"
            /protein_id="AAH09024.1"
            /db_xref="gi:14290512"
            /translation="AVLLMTIARVADGLPLASMOEDSGRDLQOYOSQAKLPRL
            NEOSPRTCTAGAMTFHYITIOGCVYLCEAAPPKLAFAYLEDLSEPEQHGK
            VPTVSRPSFIETDFDIOTKTKLYIDSRARRLGSINTELQDVQIMVANIIEVYORG
            EALSRLDSKANLSTLSKKRYODAKYLMKRSYAKLAAYAVFIMLYIVRRWML"
CDS
    BASE COUNT      347 a 291 c 344 g 357 t
ORIGIN

Query Match      13.7%; Score 21; DB 10; Length 1339;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 TAAAGACCTTGCAAGAAAGA 131
      |||||||
Db 840 TAAAGACCTTGCAAGAAAGA 860

RESULT 10
LOCUS      BC001364      1491 bp      mRNA      linear      PRI 12-JUL-2001
DEFINITION Homo sapiens, SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1, clone MGC:1960 IMAGE:3051087, mRNA, complete
            cds.

```

ACCESSION BC001364  
VERSION BC001364.1 GI:12655032  
KEYWORDS MGC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1491)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DMP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Beutlein, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantrifop, S., Thomas, P.J., Thompson, E.E., Touchman, J.W., Tsurgoun, C., Vogt, U.L., Walker, M.A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL plate 4 Row: C Column: 13  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 335139.  
FEATURES  
source  
1..1491  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:1960 IMAGE:3051087"  
/tissue\_type="Skin, melanotic melanoma."  
/clone\_1lb="NIR-MGC\_20"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
129..776  
/codon\_start=1  
/product="SEC22, vesicle trafficking protein (S. cerevisiae)-like 1"  
/protein\_id="AAH01364.1"  
/db\_xref="GI:12655033"  
/translation="MVLMTARVADQLPLASMOEDSGRDLDQYOSAKOLFRL NEOSTPCTLEAGAMTFYIEQGVCLVLCGAAPKPLAFVLELHSEFEDHGKRPVTSRPSYIEFTDIQKTKLIDSRARRLGSINTEODVORIMVNIIEVYLGK FALALDSKNNLSISKRYRODKYINMRSTYKLAIVAFIMLVYVRFMWL"  
BASE COUNT 428 a 289 c 342 g 432 t  
ORIGIN  
Query Match 13.7%; Score 21; DB 9; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AL359758  
LOCUS AL359758  
DEFINITION Human DNA sequence from clone RP11-544024 on chromosome 1, complete sequence.  
ACCESSION AL359758  
VERSION AL359758.14 GI:15620621  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 85836)  
AUTHORS Lovell, J.  
TITLE Direct Submission  
JOURNAL Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
Requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Sep 14, 2001 this sequence version replaced gi:13897376.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their respective databases: Em., EMBL, S., SWISSPROT, Tr., TrEMBL, Wp., WormPeP, information on the WormPeP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch1>  
RP11-544024 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm>  
VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone RP11-544024 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true left end of clone RP11-544024 is at 1 in this sequence. The true left end of clone RP4-646P11 is at 85737 in this sequence.  
FEATURES  
source  
1..85836  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-544024"  
/clone\_1lb="RPC1-11.2"  
30369..30381  
/note="Single clone region. Assembly confirmed by restriction digest data."  
BASE COUNT 25942 a 16986 c 16876 g 26032 t  
ORIGIN  
Query Match 13.7%; Score 21; DB 9; Length 85836;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AL391556  
LOCUS AL391556 95178 bp DNA linear HTG 10-JUL-2001  
DEFINITION Homo sapiens chromosome 1 clone RP4-630J2 map q21.1-21.3, \*\*\*  
ACCESSION SEQUENCING IN PROGRESS \*\*\*, 19 unordered pieces.  
AL391556  
VERSION AL391556.4 GI:10716368  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 95178)  
REFERENCE Direct Submission  
AUTHORS Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
TITLE CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk UK Clone  
JOURNAL requests: clonerequest@sanger.ac.uk  
COMMENT On Oct 7, 2000 this sequence version replaced gi:10040155.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: du630J2  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 56% of reads  
Chemistry: Dye-terminator ET-amersham; 43% of reads  
Consensus quality: 8614 bases at least Q40  
Consensus quality: 89533 bases at least Q30  
Consensus quality: 91633 bases at least Q20  
Insert size: 93378; sum-of-contigs  
Insert size: 125478; 4.5% error; agarose-fp  
Quality coverage: 3.24x in Q20 bases; sum-of-contigs Quality  
Coverage: 2.64x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 3063: contig of 3063 bp in length  
\* 3064 3163: gap of 100 bp  
\* 3164 6176: contig of 3013 bp in length  
\* 6177 6276: gap of 100 bp  
\* 6277 9333: contig of 3057 bp in length  
\* 9334 9433: gap of 100 bp  
\* 9434 18624: contig of 9191 bp in length  
\* 18625 18724: gap of 100 bp  
\* 18725 26968: contig of 8244 bp in length  
\* 26969 27068: gap of 100 bp  
\* 27069 37152: contig of 10084 bp in length  
\* 37153 37252: gap of 100 bp  
\* 37253 44300: contig of 7048 bp in length  
\* 44301 44400: gap of 100 bp  
\* 44401 46867: contig of 2467 bp in length  
\* 46868 46967: gap of 100 bp  
\* 46968 50594: contig of 3627 bp in length  
\* 50595 50694: gap of 100 bp  
\* 50695 53352: contig of 2658 bp in length  
\* 53353 53452: gap of 100 bp  
\* 53453 61187: contig of 7735 bp in length  
\* 61188 61287: gap of 100 bp  
\* 61288 65069: contig of 3782 bp in length  
\* 65070 65169: gap of 100 bp  
\* 65170 69334: contig of 4165 bp in length  
\* 69335 69434: gap of 100 bp  
\* 69435 77503: contig of 8069 bp in length

77504 77603: gap of 100 bp  
\* 77604 79864: contig of 2261 bp in length  
\* 79865 79964: gap of 100 bp  
\* 79965 84018: contig of 4054 bp in length  
\* 84019 84118: gap of 100 bp  
\* 84119 86596: contig of 2478 bp in length  
\* 86597 86696: gap of 100 bp  
\* 86697 92135: contig of 5439 bp in length  
\* 92136 92235: gap of 100 bp  
\* 92236 95178: contig of 2943 bp in length.  
FEATURES  
source  
1. 95178  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q21.1-21.3"  
/clone="RP4-630J2"  
/clone\_1lb="RP4-4"  
1. 3063  
/note="assembly\_fragment:00100  
fragment\_chain:1"  
3164. 6176  
/note="assembly\_fragment:00664  
fragment\_chain:1"  
6277. 9333  
/note="assembly\_fragment:00660  
fragment\_chain:1"  
9434. 18624  
/note="assembly\_fragment:00535  
fragment\_chain:2"  
18725. 26968  
/note="assembly\_fragment:00354  
fragment\_chain:2"  
27069. 37152  
/note="assembly\_fragment:00346  
fragment\_chain:2"  
37253. 44300  
/note="assembly\_fragment:00798  
fragment\_chain:3"  
44401. 46867  
/note="assembly\_fragment:00892  
fragment\_chain:3"  
46968. 50594  
/note="assembly\_fragment:00511  
fragment\_chain:3"  
50695. 53352  
/note="assembly\_fragment:00999  
fragment\_chain:4"  
53453. 61187  
/note="assembly\_fragment:01015  
fragment\_chain:4"  
61288. 65069  
/note="assembly\_fragment:00023"  
65170. 69334  
/note="assembly\_fragment:00187"  
69435. 77503  
/note="assembly\_fragment:00323"  
77604. 79864  
/note="assembly\_fragment:00464"  
79965. 84018  
/note="assembly\_fragment:00514"  
84119. 86596  
/note="assembly\_fragment:00677"  
86697. 92135  
/note="assembly\_fragment:00889"  
92236. 95178  
/note="assembly\_fragment:00941"  
BASE COUNT 27407 a 18985 c 18958 g 28022 t 1806 others  
ORIGIN  
Query Match 13.7%; Score 21; DB 2; Length 95178;  
Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 TAAAGACCTTGCAGAAAGA 131  
|||||  
Db 59055 TAAAGACCTTGCAGAAAGA 59075

RESULT 13  
AC016280 144979 bp DNA linear HTG 13-JUL-2000  
LOCUS Homo sapiens clone RP11-20N19, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC016280  
AC016280.3 GI:9123976  
VERSION  
KEYWORDS HTG: HTGS\_PHASE0.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 144979)  
Britten, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-20N19  
2 (bases 1 to 144979)  
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baldwin, J., Bara, N., Beckert, R., Boguslavsky, L., Bouhassir, B.,  
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
Cook, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,  
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,  
Galligan, J., Gardyna, S., Grant, G., Hagos, B., Harford, A., Horton, L.,  
Howard, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Lewoczky, J., Lien, C., Locke, K., Macdonald, P., Marquis, N.,  
McEwen, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,  
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6730879.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: 13944  
Center clone name: 20\_N19  
-----  
\* NOTE: This record contains 138 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1  
\* 872 971: gap of 100 bp in length  
\* 972 1881: contig of 910 bp in length  
\* 1882 1981: gap of 100 bp in length  
\* 1982 2869: contig of 888 bp in length  
\* 2870 2969: gap of 100 bp in length  
\* 2970 3880: contig of 911 bp in length  
\* 3881 3980: gap of 100 bp in length  
\* 3981 4853: contig of 873 bp in length

4854 4953: gap of 100 bp  
4954 5827: contig of 874 bp in length  
5828 5927: gap of 100 bp  
5928 6786: contig of 859 bp in length  
6787 6886: gap of 100 bp  
6887 7820: contig of 934 bp in length  
7821 7920: gap of 100 bp  
7921 8850: contig of 930 bp in length  
8851 8950: gap of 100 bp  
8951 9824: contig of 874 bp in length  
9825 9924: gap of 100 bp  
9925 10845: contig of 921 bp in length  
10846 10945: gap of 100 bp  
10946 11851: contig of 906 bp in length  
11852 11951: gap of 100 bp  
11952 12829: contig of 878 bp in length  
12830 12929: gap of 100 bp  
12930 13796: contig of 867 bp in length  
13797 13865: gap of 100 bp  
13867 14796: contig of 900 bp in length  
14797 14896: gap of 100 bp  
14897 15813: contig of 917 bp in length  
15814 15913: gap of 100 bp  
15914 16813: contig of 902 bp in length  
16816 16913: gap of 100 bp  
16916 17827: contig of 912 bp in length  
17828 17927: gap of 100 bp  
17928 18809: contig of 882 bp in length  
18810 18909: gap of 100 bp  
18910 19788: contig of 879 bp in length  
19789 19888: gap of 100 bp  
19889 20799: contig of 911 bp in length  
20800 20899: gap of 100 bp  
20900 21829: contig of 930 bp in length  
21830 21929: gap of 100 bp  
21930 22837: contig of 908 bp in length  
22838 22937: gap of 100 bp  
22938 23812: contig of 875 bp in length  
23813 23912: gap of 100 bp  
23913 24797: contig of 885 bp in length  
24798 24897: gap of 100 bp  
24898 25757: contig of 860 bp in length  
25758 25857: gap of 100 bp  
25858 26754: contig of 897 bp in length  
26755 26854: gap of 100 bp  
26855 27766: contig of 912 bp in length  
27767 27866: gap of 100 bp  
27867 28790: contig of 924 bp in length  
28791 28890: gap of 100 bp  
28891 29789: contig of 899 bp in length  
29780 29889: gap of 100 bp  
29890 30786: contig of 897 bp in length  
30787 30886: gap of 100 bp  
30887 31757: contig of 871 bp in length  
31758 31857: gap of 100 bp  
31858 32736: contig of 879 bp in length  
32737 32836: gap of 100 bp  
32837 33691: contig of 855 bp in length  
33692 33791: gap of 100 bp  
33792 34669: contig of 878 bp in length  
34670 34769: gap of 100 bp  
34770 35679: contig of 910 bp in length  
35680 35779: gap of 100 bp  
35780 36669: contig of 890 bp in length  
36670 36769: gap of 100 bp  
36770 37674: contig of 905 bp in length  
37675 37774: gap of 100 bp  
37775 38676: contig of 902 bp in length  
38677 38776: gap of 100 bp  
38777 39679: contig of 903 bp in length  
39680 39779: gap of 100 bp  
39780 40654: contig of 875 bp in length  
40655 40754: gap of 100 bp

*	40755	41619:	contig of 865 bp	in length
*	41620	41719:	gap of 100 bp	
*	41720	42559:	contig of 840 bp	in length
*	42560	42659:	gap of 100 bp	
*	42660	43575:	contig of 916 bp	in length
*	43576	43675:	gap of 100 bp	
*	43676	44570:	contig of 895 bp	in length
*	44571	44670:	gap of 100 bp	
*	44671	45577:	contig of 907 bp	in length
*	45578	45677:	gap of 100 bp	
*	45678	46590:	contig of 913 bp	in length
*	46591	46690:	gap of 100 bp	
*	46691	47556:	contig of 866 bp	in length
*	47557	47656:	gap of 100 bp	
*	47657	48546:	contig of 890 bp	in length
*	48547	48646:	gap of 100 bp	
*	48647	49517:	contig of 871 bp	in length
*	49518	49617:	gap of 100 bp	
*	49618	50511:	contig of 894 bp	in length
*	50512	50611:	gap of 100 bp	
*	50612	51543:	contig of 932 bp	in length
*	51544	51643:	gap of 100 bp	
*	51644	52559:	contig of 916 bp	in length
*	52560	52659:	gap of 100 bp	
*	52660	53547:	contig of 888 bp	in length
*	53548	53647:	gap of 100 bp	
*	53648	54528:	contig of 881 bp	in length
*	54529	54628:	gap of 100 bp	
*	54629	55500:	contig of 872 bp	in length
*	55501	55600:	gap of 100 bp	
*	55601	56503:	contig of 903 bp	in length
*	56504	56603:	gap of 100 bp	
*	56604	57517:	contig of 914 bp	in length
*	57518	57617:	gap of 100 bp	
*	57618	58500:	contig of 883 bp	in length
*	58501	58600:	gap of 100 bp	
*	58601	59484:	contig of 884 bp	in length
*	59485	59584:	gap of 100 bp	
*	59585	60487:	contig of 903 bp	in length
*	60488	60587:	gap of 100 bp	
*	60588	61508:	contig of 921 bp	in length
*	61509	61608:	gap of 100 bp	
*	61609	62533:	contig of 925 bp	in length
*	62534	62633:	gap of 100 bp	
*	62634	63476:	contig of 843 bp	in length
*	63477	63576:	gap of 100 bp	
*	63577	64458:	contig of 882 bp	in length
*	64459	64558:	gap of 100 bp	
*	64559	65442:	contig of 884 bp	in length
*	65443	65542:	gap of 100 bp	
*	65543	66455:	contig of 913 bp	in length
*	66456	66555:	gap of 100 bp	
*	66556	67462:	contig of 907 bp	in length
*	67463	67562:	gap of 100 bp	
*	67563	68550:	contig of 988 bp	in length
*	68551	68650:	gap of 100 bp	
*	68651	69609:	contig of 959 bp	in length
*	69610	69709:	gap of 100 bp	
*	69710	70779:	contig of 1070 bp	in length
*	70780	70879:	gap of 100 bp	
*	70880	71850:	contig of 971 bp	in length
*	71851	71950:	gap of 100 bp	
*	71951	72978:	contig of 1028 bp	in length

Query Match 13.7%: Score 21; DB 2: Length 144979;  
Best Local Similarity 100.0%; Pred No. 1,8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 112 AAAGACCTTGCAAGAAAGAA 132  
|||||  
DB 136773 AAAGACCTTGCAAGAAAGAA 136793

RESULT 14  
AC017081 149462 bp DNA linear PRI 09-JAN-2002  
LOCUS  
AC017081 Homo sapiens BAC clone RP11-470J24 from 2, complete sequence.  
AC017081  
AC017081.8 GI:18093316  
KEYWORDS  
HTG.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 149462)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
2 (bases 1 to 149462)  
Nguyen, C., Doebber, A. and Kozlowski, A.  
The sequence of Homo sapiens BAC clone RP11-470J24  
Unpublished (2001)  
3 (bases 1 to 149462)  
Waterston, R.H.  
Direct Submission  
Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 149462)  
Waterston, R.  
Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 9, 2002 this sequence version replaced gi:14165368.  
-----  
Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [saplensew@wustl.edu](mailto:saplensew@wustl.edu)  
-----  
Summary Statistics  
Center project name: H\_NH0470J24  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBAC3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-310K15, 2000 bp overlap; the clone sequenced to the right is RP11-156A1, 2000 bp overlap.

Actual start of this clone is at base position 190775 of  
RP11-310K15; actual end is at base position 28935 of RP11-156A1.

The sequence between 81601 and 81703 is covered only by a PCR  
product of clone DNA. Unresolved tandem repeat regions exist  
between 81548 and 83183, 126196 and 127426.

## FEATURES

```
source
1..149462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-470J24"
/clone_1lb="RPC1-11"
141..282
/rpt_family="MIR"
repeat_region
581..621
/rpt_family="(TGCG)n"
repeat_region
965..1037
/rpt_family="GC-rich"
misc_feature
1127..1561
/note="match to EST AA056210 (NID:91548612) zf62906.s1"
misc_feature
1771..2127
/note="match to EST AW614820 (NID:97320006) hg80c08.x1"
repeat_region
2035..2195
/rpt_family="MER1_type"
repeat_region
2216..2419
/rpt_family="MER1_type"
repeat_region
2604..2844
/rpt_family="MIR"
misc_feature
2991..3324
/note="match to EST R35058 (NID:9791959) yh86d09.r1"
misc_feature
3207..3758
/note="match to EST AW452004 (NID:96992780)"
repeat_region
3486..3527
/rpt_family="AT-rich"
misc_feature
3501..3745
/note="match to EST AA054208 (NID:91545131) zf51903.r1"
repeat_region
3981..4017
/rpt_family="T-rich"
repeat_region
4231..4273
/rpt_family="AT-rich"
misc_feature
5518..5847
/note="match to EST R80994 (NID:9857275) y194c08.r1"
repeat_region
5606..5633
/rpt_family="AT-rich"
repeat_region
5722..5758
/rpt_family="AT-rich"
repeat_region
5802..5831
/rpt_family="(TTTAA)n"
repeat_region
5808..6116
/rpt_family="Alu"
repeat_region
6133..6244
/rpt_family="MER1_type"
repeat_region
6965..7000
/rpt_family="AT-rich"
repeat_region
7020..7094
/rpt_family="MER1_type"
repeat_region
7982..8006
/rpt_family="AT-rich"
repeat_region
8000..8436
/rpt_family="L2"
repeat_region
8437..8540
/rpt_family="L1"
repeat_region
8541..8696
/rpt_family="L2"
repeat_region
8722..9684
/rpt_family="L1"
repeat_region
9378..9422
/rpt_family="(TTG)n"
repeat_region
9685..9973
/rpt_family="Alu"
repeat_region
9974..10197
```

```
repeat_region /rpt_family="L1"
10280..10305 /rpt_family="(TATG)n"
repeat_region /rpt_family="AT-rich"
10320..10341 /rpt_family="L1"
10333..10679 /rpt_family="L1"
10681..10757 /rpt_family="L1"
10758..11039 /rpt_family="Alu"
11040..12011 /rpt_family="L1"
11999..12591 /rpt_family="L1"
12592..12706 /rpt_family="Alu"
12707..12952 /rpt_family="L1"
12966..13140 /rpt_family="L1"
13139..16061 /rpt_family="L1"
16063..17107 /rpt_family="L1"
17108..17161 /rpt_family="L1"
17163..17468 /rpt_family="MER1_type"
17471..17675 /rpt_family="MER1_type"
17677..17835 /rpt_family="MER1_type"
17909..18144 /rpt_family="L1"
18177..18501 /rpt_family="L1"
18519..20293 /rpt_family="L1"
20490..20906 /rpt_family="L1"
21258..21365 /rpt_family="L1"
21353..21393 /rpt_family="MIR"
21366..21684 /rpt_family="(T)n"
21543..21566 /rpt_family="Alu"
21685..21819 /rpt_family="AT-rich"
22160..22640 /rpt_family="MIR"
22878..23228 /rpt_family="MER1_type"
23687..23991 /rpt_family="L1"
repeat_region
```

Query Match 13.7%; Score 21; DB 9; length 149462;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AAAGACCTGCAAGAAAACAA 132  
DB 111213 AAAGACCTGCAAGAAAACAA 111233

RESULT 15  
AC024119 AC024119 150436 bp DNA linear HTG 12-OCT-2000  
LOCUS Homo sapiens chromosome 3, WORKING DRAFT SEQUENCE, 22 unordered  
DEFINITION pieces  
AC024119  
AC024119.2 GI:10305055

KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 150436)  
Xu,S., Zhao,Y., Lin,W., Dong,H., Wan,M., Zhang,C., Gu,W., Tu,Y.,  
Jia,J., Wu,C., Lu,G., Zhong,M., Zhou,Y., Ren,S., Fu,G., Chen,Z. and  
Huang,M.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2000) Genomic Dept., Chinese National Human  
Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,  
Shanghai 201203, P. R. China  
On Sep 26, 2000 this sequence version replaced gi:7025692.  
COMMENT -----Genome Center-----  
Center:Beijing  
Center code:Beijing  
Website:http://hgclgt.ac.cn  
http://www.genomics.org.cn  
Contact:hgclgt.ac.cn  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
2766: contig of 2766 bp in length  
2767 gap of unknown length  
5770: contig of 3004 bp in length  
5771 gap of unknown length  
8549: contig of 2779 bp in length  
8550 gap of unknown length  
10693: contig of 2144 bp in length  
10694 gap of unknown length  
13973: contig of 3280 bp in length  
13974 gap of unknown length  
16980: contig of 3007 bp in length  
16981 gap of unknown length  
20419: contig of 3439 bp in length  
20420 gap of unknown length  
23275: contig of 2856 bp in length  
23276 gap of unknown length  
26837: contig of 3562 bp in length  
26838 gap of unknown length  
31026: contig of 4189 bp in length  
31027 gap of unknown length  
35204: contig of 4178 bp in length  
35205 gap of unknown length  
42487: contig of 7283 bp in length  
42488 gap of unknown length  
49419: contig of 6932 bp in length  
49420 gap of unknown length  
57524: contig of 8105 bp in length  
57525 gap of unknown length  
65794: contig of 8270 bp in length  
65795 gap of unknown length  
69733: contig of 3939 bp in length  
69734 gap of unknown length  
81835: contig of 12102 bp in length  
81836 gap of unknown length  
87228: contig of 5393 bp in length  
87229 gap of unknown length  
106006: contig of 18778 bp in length  
106007 gap of unknown length  
122564: contig of 16558 bp in length  
122565 gap of unknown length  
128482: contig of 5918 bp in length  
128483 gap of unknown length  
150436: contig of 21954 bp in length.  
FEATURES  
Location/Qualifiers  
1..150436

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone\_lib="542323"  
BASE COUNT 45301 a 29770 c 31151 g 44136 t 78 others  
ORIGIN  
Query Match 13.7%; Score 21; DB 2; length 150436;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 111 TAAAGACCTTGCAAGAAAGA 131  
|||||  
Db 91838 TAAAGACCTTGCAAGAAAGA 91858

Search completed: August 3, 2002, 01:43:18  
Job time: 40768 sec



Mon Aug 5 11:51:53 2002

us-08-973-363-5.011.rge

Page 11



**THIS PAGE BLANK (USPTO)**



PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 DR P-PSDB: AAM08149.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
 CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
 CC sequences of this region are provided in AAM08146-49. The CHD-1A  
 CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 S0 Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;  
 XX

Query Match	90.2%	Score 138:	DB 18:	Length 153;
Best Local Similarity	100.0%	Pred. NC.	3.6e-60;	
Matches 138; Conservative	0;	Mismatches	0;	Gaps 0;

Oy	16	CCAGGTAAGAAACACACAGGCAAGACAGTTCGACGACCAAGAACCAACAGGCAACAGCACTTG	75
Oy	16	CCAGGTAAGAAACACACAGGCAAGACAGTTCGACGACCAAGAACCAACAGGCAACAGCACTTG	75
Db	16	ccagtaagaacacccacgacgaagcagctgtgcagcccaagaacccacacgacgaagcagctgtg	75
Oy	76	CAGACCCGCTGGAGATTAACCTCATTTAAATTAAGTGAATTAAGACCTTGGCAAGAAAGAAAGTG	139
Db	76	cagaccctgtcagatctaacctcaatcaatctactgataaagacctctgcagaagaagaagctg	135
Oy	136	CAAGACTTACTGCTGCA	153
Db	136	caaaagacttactgctgtcga	153

```

RESULT      2
AAT42758
ID  AAT42758 standard; DNA; 153 BP.

```

AC AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

Bird; sex determination; 'chromodomain-Helicase-DNA binding 1';

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

	Location/Qualifiers
FH	Key
FT	misc_difference 52..81

FT /note="bases 52-81 are a repeat of bases 22-51  
FT and are ignored in the translated amino  
FT acid sequence given in Fig 3"  
...

PN W09639505-A1.

PD 12-DEC-1996

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiward B;

DR WPI: 1997-043127/04.  
DR P-PSDB; AAM08148.  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determ. and to control sex of progeny  
PT  
XX  
PS Claim 8; Fig 3; 76pp: English.  
XX  
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),  
CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and  
CC and the great tit CHD-W gene (AA142759). Translated amino acid  
CC sequences of this region are provided in AAM08146-49. The CHD-1A  
CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine  
CC sex in birds and can be used to identify the sex of an embryo,  
CC foetus etc. and to manipulate the sex of progeny.  
CC  
SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match	34.6%;	Score 53;	DB 18;	Length 153;
Best Local Similarity	100.0%;	Pred. No. 2.8e-17;		
Matches 53; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

QY 76 CAGACCCGTGCAGATTACCTTAATTAATCTGAATATAAGACCCTTGCAAGAAA 128  
|||||  
76 cagaccctgcagattaccttaataatctgaataaagacccttgcaagaaa 128  
Db

RESULT	3
AAT42754	
ID	AAT42754 standard; cDNA; 1316 BP

AC AAT42754;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-W gene (partial sequence).

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian,

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN W09639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in birds - used for sex determin. and to control sex of progeny

PS Claim 1; Fig 8; 76pp; English.

CC The chicken CHD-W gene (AA1742754) acting alone or in conjunction with  
CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
CC female development in birds. The sequence of CHD-W was deduced  
CC from 2 clones isolated from a 10-day chick embryo library using  
CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is  
CC located on the W chromosome. Probes based on CHD-W and CHD-1A give  
CC a W chromosome-specific signal on hybridisation to genomic DNA of a  
CC non-ratite bird and can be used for sex determin. of a bird. CHD-W  
CC nucleic acids can also be used to control the sex of progeny of a  
CC bird.

```

XX      SQ      Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;
SQ      Query Match
          30.7%; Score 47; DB 18; Length 1316;
          Best Local Similarity 100.0%; Pred. No. 2,9e-14;
          Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      82 CGTCGAGATTACCTGATTAATTAAGACCTTGCAAGAAA 128
          |||||||
Db      1189 cgtgagattacctctaatctaatctgataaagaccttgcaagaaa 1235

RESULT 4
AAAT42757
ID      AAAT42757 standard; DNA; 153 BP.
AC      AAAT42757;
XX      12-MAR-1997 (first entry)
XX      Chick CHD-1A gene fragment.
XX      DE
XX      Bird: sex determination; chromodomain-Helicase-DNA binding 1;
XX      KM CHD-1A; CHD-W; W chromosome; ss.
XX      OS
XX      Gallus sp.
XX      Key Location/Qualifiers
XX      FT misc_difference 52..81
XX      FT /*tag= a
XX      FT /*note= "bases 52-81 are a repeat of bases 22-51
XX      FT and are ignored in the translated amino
XX      FT acid sequence given in Fig 3"
XX      PN MO9639505-A1.
XX      PD 12-DEC-1996.
XX      PF 05-JUN-1996; 96MO-GB01341.
XX      PR 06-JUN-1995; 95GB-0011439.
XX      PA (ISIS-) ISIS INNOVATION LTD.
XX      PI Griffiths R, Tiwari B;
XX      PI Griffiths R, Tiwari B;
XX      DR WPI: 1997-043127/04.
XX      DR P-PSDB: AAM08147.
XX      PT Avian chromodomain-helicase-DNA binding genes determine sex in
XX      PT birds - used for sex determin. and to control sex of progeny
XX      PS Claim 8; Fig 3; 76pp; English.
XX      CC Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
XX      CC to portions of the chicken CHD-1A (A - Avian) gene (AAAT42757),
XX      CC chicken CHD-W (W refers to the W chromosome) gene (see also AAAT42758)
XX      CC and the great tit CHD-W gene (see also AAAT42759). Translated amino
XX      CC acid sequences of this region are provided in AAM08146-49. The
XX      CC CHD-1A (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes
XX      CC determine sex in birds and can be used to identify the sex of an
XX      CC embryo, fetus etc. and to manipulate the sex of progeny.
XX      SO Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match
          24.8%; Score 38; DB 18; Length 153;
          Best Local Similarity 100.0%; Pred. No. 1e-09;
          Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      91 TACCTCATTAATTAAGACCTTGCAAGAAA 128
          |||||||
XX      21-OCT-1998 (first entry)

Db      91 taccctcatctaataatgaataaagaccttgcaagaaa 128

RESULT 5
AAAT42751
ID      AAAT42751 standard; cDNA; 6608 BP.
AC      AAAT42751;
XX      12-MAR-1997 (first entry)
XX      DE
XX      Chicken CHD-1A gene.
XX      KM Bird: sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX      KM CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX      OS
XX      Gallus sp.
XX      Key Location/Qualifiers
XX      FT CDS 228..5390
XX      FT /*tag= a
XX      PN MO9639505-A1.
XX      PD 12-DEC-1996.
XX      PF 05-JUN-1996; 96MO-GB01341.
XX      PR 06-JUN-1995; 95GB-0011439.
XX      PA (ISIS-) ISIS INNOVATION LTD.
XX      PI Griffiths R, Tiwari B;
XX      PI Griffiths R, Tiwari B;
XX      DR WPI: 1997-043127/04.
XX      PT Avian chromodomain-helicase-DNA binding genes determine sex in
XX      PT birds - used for sex determin. and to control sex of progeny
XX      PS Claim 1; Fig 5; 76pp; English.
XX      CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with
XX      CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate
XX      CC female development in birds. The sequence of CHD-1A was deduced
XX      CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA
XX      CC library using a great tit CHD-W sequence (see also AAAT42755) as probe.
XX      CC The CHD-1A (A - Avian) gene shows close identity to the mouse CHD-1
XX      CC gene (see also AAAT42756-57). It is located on an autosome or 2
XX      CC chromosome. Probes based on CHD-W and CHD-1A give a non-rare
XX      CC specific signal on hybridisation to genomic DNA of a non-rare
XX      CC bird and can be used for sex determin. of a bird. CHD-1A nucleic
XX      CC acids can also be used to control the sex of the progeny of a bird.
XX      SO Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match
          24.8%; Score 38; DB 18; Length 6608;
          Best Local Similarity 100.0%; Pred. No. 1e-09;
          Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      91 TACCTCATTAATTAAGACCTTGCAAGAAA 128
          |||||||
Db      4140 taccctcatctaataatgaataaagaccttgcaagaaa 4177

RESULT 6
AAV49574
ID      AAV49574 standard; cDNA to mRNA; 1409 BP.
AC      AAV49574;
XX      21-OCT-1998 (first entry)
XX      DT

```

DE	Human lymphoma cell line U937 clone HP10136 cDNA #1.
KX	Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW	differentiation; immune system; stimulator; suppressor; regulator;
KV	hematopoiesis; activin; inhibitor; chemotactic; chemoketic; receptor;
KX	haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX	
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	CDS 82..729
FT	/*tag=^
FT	/product= "transmembrane domain containing protein"
XX	
PN	MO9821328-AZ.
PD	22-MAY-1998.
XX	
PE	07-NOV-1997; 97WO-JP04056.
XX	
PR	13-NOV-1996; 96JP-0301429.
XX	
PA	(PROT-) PROTEGENE INC.
XX	(SAGA ) SAGAMI CHEM RES CENTRE.
PI	Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR	MPJ: 1998-297932/26.
DR	P-PSDB; AAM64546.
PT	Human protein having transmembrane domain - useful for, e.g.
PT	research and nutrition
PS	Claim 4; Page 153-155; 205pp; English.
CC	AAV49550-V49559 are cDNA sequences which encode human proteins containing
CC	a transmembrane domain. These proteins can be used for, e.g. research
CC	and nutrition, and may have cytokine and cell
CC	proliferation/differentiation, immune stimulating/suppressing,
CC	haematopoesis regulating, tissue growth, activin/inhibin,
CC	chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/Ligand,
CC	anti-inflammatory or tumour inhibition activity.
XX	
SQ	Sequence 1409 BP; 377 A; 278 C; 325 G; 429 T; 0 other;
Query Match	13.7%; Score 21; DB 19; Length 1409;
Best Local Similarity	100.0%; Pred. No. 0.38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	111 TAAAGACCTTGCAAGAAAACA 131
Db	842 taaagaccttgcaagaataa 862
RESULT 7	
AAX56240	
ID	AAX56240 standard; CDNA; 1462 BP.
XX	
AAX56240;	
XX	
DT	16-JUL-1999 (first entry)
XX	
DE	Human CBFBBA01 vesicle trafficking protein SEC22b gene.
XX	
KW	Human; vesicle trafficking protein; SEC22b; CBFBBA01; cancer;
KV	autoimmune disease; diabetes mellitus; multiple sclerosis;
KW	diagnosis; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09921988-A1.
XX	

PD	06-MAY-1999.
XX	
PF	29-OCT-1997; 97WO-CN00115.
XX	
PR	29-OCT-1997; 97WO-CN00115.
XX	
PA	(USSH-) UNITV SHANGHAI SECOND MEDICAL.
PX	
PI	Wang Y, Wu J;
DR	WPI: 1999-312961/26.
xx	P-PSDB; AAY09512.
xx	
PT	CBFBA01 protein useful in treating cancer, autoimmune disease,
xx	diabetes mellitus and multiple sclerosis
PS	Claim 2; Page 21; 29pp; English.
xx	
CC	The present sequence encodes CBFBA01, which is a protein homologous to
CC	murine vesicle trafficking protein sec22b. CBFBA01 is a
CC	cycloplasmically-orientated integral membrane protein, located in the
CC	endoplasmic reticulum and golgi membranes. CBFBA01 polynucleotide (I),
CC	vectors containing (I) and recombinant host cells are useful for
CC	recombinant production of CBFBA01. Both (I), CBFBA01 and antibodies
CC	against CBFBA01 are useful as research reagents, for screening assays
CC	and in diagnostic assays, especially for cancer, autoimmune disease,
CC	diabetes mellitus and multiple sclerosis. Antagonists and agonists of
CC	CBFBA01 can be used to inhibit or enhance, respectively, the activity
CC	of CBFBA01 or expression of (I). Anti-CBFBA01 antibodies and CBFBA01
CC	or its fragments can be used in vaccines.
SQ	Sequence 1462 BP; 419 A; 286 C; 323 G; 434 T; 0 other;
Query Match	13.7%; Score 21; DB 20; Length 1462;
Best Local Similarity	100.0%; Pred. NO. 0.38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	111 TAAAGACCTTGCAAGAAAAGA 131 
Db	826 taaagaccttgcaagaagaaga 846
RESULT 8	
AAT11338	
ID	AAT11338 standard; cDNA; 557 BP.
XX	
AC	AAT11338;
XX	
DT	05-SEP-1996 (first entry)
XX	
DE	Aft-1 interacting protein (partial ascorbate peroxidase) cDNA.
XX	
KM	Arabidopsis fourteen-three-three 1; AFT1; Cruciferae; regulation;
KW	transcription activator; gene expression modulator; crucifer;
RW	signal transduction; defence response; gene therapy; transgenic;
plant; ss.	
XX	
OS	Arabidopsis sp.
XX	
FH	Key Location/Qualifiers
FT	CDS 2..370
FT	/tag= a
FT	/product= partial ascorbate peroxidase
FT	/note= "plant defence related protein"
EP693554-A1.	
PN	
XX	
PD	24-JAN-1996.
XX	
PE	22-JUN-1995; 95EP-0109669.
XX	
PR	23-JUN-1994; 94US-0266451.

XX (GENO ) GEN HOSPITAL CORP.  
PA (MASS ) MASSACHUSETTS GEN HOSPITAL.  
XX Goodman HM, Zhang H;  
XX WPI; 1996-078601/09.  
DR P-PSDB; AAR89746.  
XX  
PT Recombinant transcriptional activator APT-1 - isolated from  
PT Arabidopsis and useful for modifying gene expression and modulating  
PT plant defence responses to pathogens  
XX  
PS Disclosure; Page 21; 47pp; English.  
XX  
CC AAT11338 encodes a partial ascorbate peroxidase, a plant defence-  
CC related protein (PDRP). AAT11338 and other cDNA molecules encoding  
CC PDRPs were isolated using a LexA/AP1 33-194 fusion protein, i.e.  
CC APT1 (Arabidopsis 14-3-3-1) protein residues 33-194 fused to LexA,  
CC a DNA binding protein, in an interaction trap assay. APT1 is a  
CC transcriptional activator protein and may be used to enhance,  
CC control or modify plant gene expression, e.g. in the regulation of  
CC plant storage components (esp. napin, legumin or phaseolin). APT1  
CC can also be used to modulate signal transduction events involved  
CC in plant defence responses to pathogens such as fungi, nematodes,  
CC bacteria and viruses.  
XX  
SQ Sequence 557 BP; 164 A; 110 C; 133 G; 150 T; 0 other;  
  
Query Match 13.1%; Score 20; DB 17; Length 557;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 31 CAGGCAAGCAGTTCGACGAC 50  
|||||  
Db 255 caggaagaagcagctgcagac 274  
  
RESULT 9  
AAC47521  
ID AAC47521 standard; DNA; 1157 BP.  
XX  
AC AAC47521;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54126.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX  
XX PD 06-SEP-2000.  
XX  
XX  
XX PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0128843.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132853.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134770.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144065.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145513.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147792.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 13.1%; Score 20; DB 21; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 31 CAGGCAAGCAGTTCGAC 50  
Db 860 caggcaagcagttgcagac 879  
  
RESULT 10  
AAC34655  
ID AAC34655 standard; DNA; 1160 BP.  
XX AC AAC34655;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7404.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
XX KW protein identification; signal transduction pathway;  
XX KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.



PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-012863.  
PR 11-MAY-1999; 99US-0134286.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137232.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138004.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139432.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142380.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 20-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0154779.  
PR 23-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 13.1%; Score 20; DB 21; Length 1160;  
Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

OY 31 CAGGCAAGCAGTTGCAGAC 50  
|||||  
Db 863 cagcgaagcagctgcagac 882

RESULT 11  
AA298313  
ID AA298313 standard; DNA; 2580 BP.

AC AA298313;

DT 14-JUN-2000 (first entry)

DE A. thaliana gene involved in environmental stress tolerance.

KM Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;  
KM dehydration; drought; heat stress; salinity; osmotolerance; ds.

OS Arabidopsis thaliana.

PN WO200008187-A2.

PD 17-FEB-2000.

PF 04-AUG-1999; 99WO-EP05652.

PR 04-AUG-1998; 98EP-0202634.

PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PI Lee JH, Verbruggen N;

DR WPI: 2000-205726/18.

P-PSDB; AAT77933.

PT Isolation of polynucleic acids useful for producing transgenic plant by  
isolating genes involved in tolerance to environmental stress  
PS Claim 4; Page 112-115; 312pp; English.

CC The invention relates to isolation of coding sequences and/or genes  
CC involved in tolerance to environmental stress in plants. The sequences  
CC (AA298305-298365) are useful for producing a transgenic plant having  
CC enhanced tolerance or resistance to environmental stress conditions such  
CC as anaerobic, flooding, cold, dehydration, drought, heat stress or  
CC salinity. This is useful for producing improved yield, growth,  
CC development and productivity under environmental stress conditions, and  
CC also provides growth of crops in areas where they cannot grow without  
CC the induced osmotolerance. Sequences AA298305-365 represent  
CC polynucleotide sequences from A. thaliana that are involved in  
CC environmental stress tolerance.

Sequence 2580 BP; 677 A; 459 C; 544 G; 900 T; 0 other;

Query Match 13.1%; Score 20; DB 21; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 1.2; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

OY 31 CAGGCAAGCAGTTGCAGAC 50  
|||||  
Db 2449 cagcgaagcagctgcagac 2468

RESULT 12

AA158977  
ID AA158977 standard; cDNA; 2532 BP.

AC AA158977;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1180.

KM Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KM Leukemia; ss.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Dirmanac RT;

DR WPI: 2001-442253/47.

P-PSDB; AAM39821.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
isolating genes involved in tolerance to environmental stress  
PS Claim 1; SEQ ID NO 1180; 10078pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.

```
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2532 BP: 648 A; 654 C; 688 G; 542 T; 0 other:

Query Match
Best Local Similarity 12.48; Score 19; DB 22; Length 2532;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGATGACCCAGATAGAAA 27
1776 tgaatgaccagataagaaa 1794

RESULT 13
AAS20000
ID AAS20000 standard; DNA; 44861 BP.
XX
AC AAS20000;
XX
DT 26-MAR-2002 (first entry)
XX
DE DNA encoding pyridoxal (pyridoxine, vitamin B6) kinase (PDXK).
XX
KM Pyridoxal kinase; pyridoxine; vitamin B6; PDXK
KM autocolumne polyglutandular disease type 1; transgenic animal;
KM gene therapy; chromosome 21q22.3; ds.
XX
OS Homo sapiens.
XX
FH Key
FH variation
FT /tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT 4093..40861
FT /tag= b
FT /product= "PDXK"
FT /note= "Pyridoxal (pyridoxine, vitamin B6) kinase,
FT 4093..4179
FT /tag= c
FT /number= 1
FT 4180..18866
FT /tag= d
FT /number= 1
FT replace(4252,T)
FT /tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT 18867..18921
FT /tag= f
FT /number= 2
FT 18922..26464
FT /tag= g
FT /number= 2
FT 26465..26569
FT /tag= h
FT /number= 3
FT 26570..28521
FT /tag= i
FT /number= 3
FT 28522..28605
FT /tag= j
FT /number= 4
FT 28606..30876
FT /tag= k
FT /number= 4
FT replace(28631,G)
FT /tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28683,A)
FT /tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT 30877..30923
FT /tag= n
FT /number= 5
FT 30924..33792
FT /tag= o
FT /number= 5
FT replace(33776,C)
FT /tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT 33793..33880
FT /tag= q
FT /number= 6
FT replace(33801,A)
FT /tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT replace(33831,G)
FT /tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT replace(33840,A)
FT /tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT 33881..35300
FT /tag= u
FT /number= 6
FT /cons_splice=(5'site:NO,3'site:NO)
FT replace(33922,G)
FT /tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35202,G)
FT /tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT 35301..35343
FT /tag= x
FT /number= 7
FT 35344..37315
FT /tag= y
FT /number= 7
FT replace(35497,G)
FT /tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37287,A)
FT /tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37296,A)
FT /tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT 37316..37427
FT /tag= ac
FT /number= 8
FT 37428..38381
FT /tag= ad
FT /number= 8
FT /cons_splice=(5'site:YES,3'site:NO)
FT replace(37545,A)
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT 38382..38517
FT /tag= af
FT /number= 9
FT replace(38397,T)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT 38518..40495
FT /tag= ah
FT /number= 9
FT replace(38622,C)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT 40496..40562
FT /tag= aj
FT /number= 10
FT replace(40516,C)
FT /tag= ak
```



OY 97 ATTAAATTACTGAATAA 114  
 |||||  
 DB 492 attaatctactgaataaa 509

## RESULT 15

AAAF89129  
 ID AAF89129 standard; DNA; 1580 BP.

AC AAF89129;

DT 13-JUL-2001 (first entry)

DE Human homoglutamine-rich factor 56 coding sequence.

KW Human: homoglutamine-rich factor 56; HQRF: cancer; angiocardiopathy;  
 KW nervous system disease; immune disease; phlogosis; ds.

OS Homo sapiens.

FN Key Location/Qualifiers  
 FT 27..1547  
 CDS /tag=a  
 /product="homoglutamine-rich factor 56"

WO200127148-A1.

PD 19-APR-2001.

PF 08-OCT-2000; 2000WO-CN00310.

PR 10-OCT-1999; 99CN-0116946.

PA (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.

PI Mao Y, Xie Y;

DR WPI; 2001-281976/29.

DR P-PSDB; AAB83299.

PT Homo (human) glutamine(Q)-rich factor 56 applicable in diagnosis and  
 PT treatment of cancer, angiocardiopathy, nervous system diseases, immune  
 PT diseases and phlogosis

PS Claim 6; Page 21-22; 29pp; Chinese.

CC The present invention provides the protein and coding sequences of human  
 CC homoglutamine-rich factor 56 (HQRF). This is useful in the treatment of  
 CC cancer, angiocardiopathy, nervous system diseases, immune diseases and  
 CC phlogosis. The present sequence is the homoglutamine-rich factor 56  
 CC coding sequence.

SO Sequence 1580 BP; 463 A; 444 C; 346 G; 327 T; 0 other:

Query Match 11.88; Score 18; DB 22; Length 1580;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ATAGAAACCCACAGGCAA 37  
 |||||

DB 237 ataagaaccacagcgcaa 254

Search completed: August 3, 2002, 01:50:11  
 Job time: 32196 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:37 ; Search time 7016.61 Seconds  
(without alignments)  
294.307 Million cell updates/sec

Title: US-08-973-363-5

Sequence: 1 ATTTACTGATGATGACCCAGCA.....TGCAGAGCTTACTGCTGCA 153

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:  
1: em\_esthda:\*  
2: em\_esthmu:\*  
3: em\_estln:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_iny:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	14.4	167	10	BE842996
C 2	21	13.7	314	9	AA354813
C 3	21	13.7	322	10	BF289372
C 4	21	13.7	355	9	AI902726
C 5	21	13.7	438	9	AA234796
C 6	21	13.7	461	9	AA456385
C 7	21	13.7	474	10	BG66575
C 8	21	13.7	488	10	BG573124
C 9	21	13.7	489	9	AA64549
C 10	21	13.7	492	9	BE198540
C 11	21	13.7	515	9	AW579514
C 12	21	13.7	560	10	BM28157
C 13	21	13.7	589	9	AI412929
C 14	21	13.7	594	9	AM993220
C 15	21	13.7	597	10	BF168086
C 16	21	13.7	616	9	AM992861
C 17	21	13.7	616	10	BF690915

C 18	21	13.7	624	9	AM992921
C 19	21	13.7	628	9	AM958448
C 20	21	13.7	631	10	BM229289
C 21	21	13.7	641	9	AM992919
C 22	21	13.7	643	9	AM993201
C 23	21	13.7	647	9	AM993283
C 24	21	13.7	658	9	AM993467
C 25	21	13.7	675	9	AV232349
C 26	21	13.7	684	9	AV232349
C 27	21	13.7	694	9	AM993481
C 28	21	13.7	698	9	AM993471
C 29	21	13.7	706	9	AM993268
C 30	21	13.7	778	10	BI852172
C 31	21	13.7	793	10	BG182192
C 32	21	13.7	874	10	BG426048
C 33	21	13.7	924	10	BG180912
C 34	21	13.7	982	10	BI105781
C 35	21	13.7	1358	11	AK010363
C 36	21	13.7	1394	11	AK017237
C 37	21	13.7	1462	11	AF047442
C 38	20	13.1	294	9	AV531270
C 39	20	13.1	337	9	AM996787
C 40	20	13.1	366	9	AV566736
C 41	20	13.1	379	9	AV521243
C 42	20	13.1	411	10	Z34197
C 43	20	13.1	430	9	AL601246
C 44	20	13.1	502	9	AV518366
C 45	20	13.1	502	9	AV521265

## ALIGNMENTS

RESULT 1  
LOCUS BE842996/c  
DEFINITION MR4-ST0240-240700-013-e10 ST0240 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE842996  
VERSION BE842996.1  
KEYWORDS GI:10275374  
SOURCE EST

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 167)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Birones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldmann,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=612-MR4-ST0240-240  
700-013-e10-ft3-2000-07-24-ft4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 26  
High quality sequence stop: 167.  
Location/Qualifiers  
1..167  
/organism="Homo sapiens"

## FEATURES

source

```

/db_xref="taxon:9606"
/clone_lib="ST0240"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      38 a      42 c      29 g      58 t
ORIGIN
Query Match      14.4%; Score 22; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      111 TAAAGACCTTGCAAGAAAGA 132
        |||||||
Db       46 TAAAGACCTTGCAAGAAAGA 25

RESULT 2
AA354813      314 bp      mRNA      linear      EST 21-APR-1997
DEFINITION   Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION    AA354813
VERSION       AA354813.1 GI:2007319
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 314)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirness,E.F., Weinstein,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gneltun,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-T., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudel,D.M., Shirley,R.,
Small,R.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venner,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
TITLE
JOURNAL
MEDLINE
COMMENT
Other_ESTs: THC180658
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/ngi/ngi.html)
Seq primer: M13 Reverse.
FEATURES
source
1..314
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="ATCC (ihost):155260"
/db_xref="taxon:9606"

```

```

/clone_lib="Jurkat T-cells V"
/cell_type="T-Lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      90 a      62 c      68 g      94 t
ORIGIN
Query Match      13.7%; Score 21; DB 9; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      111 TAAAGACCTTGCAAGAAAGA 131
        |||||||
Db       61 TAAAGACCTTGCAAGAAAGA 81

RESULT 3
BF289372      322 bp      mRNA      linear      EST 28-NOV-2000
LOCUS        BF289372
DEFINITION   EST453963 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
Rattus norvegicus cDNA clone R01HE60, mRNA sequence.
ACCESSION    BF289372
VERSION       BF289372.1 GI:11220442
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 322)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizl,B., Perlea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
CONTACT: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
Location/Qualifiers
1..322
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="R01HE60"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROY, RBR, RKL, RLI, RPL, RUU, REM, RNU, RSP
, RHE, RPC, RPN"
BASE COUNT      77 a      71 c      80 g      94 t
ORIGIN
Query Match      13.7%; Score 21; DB 10; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      111 TAAAGACCTTGCAAGAAAGA 131
        |||||||
Db       18 TAAAGACCTTGCAAGAAAGA 38

RESULT 4
AI902726      355 bp      mRNA      linear      EST 30-MAR-2000
LOCUS        AI902726
DEFINITION   OV-BT011-101198-028 BT011 Homo sapiens cDNA, mRNA sequence.
ACCESSION    AI902726

```



VERSION A1902726.1 GI:6493113  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 355)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,  
Nagai,M.A., da Silva,M., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/seq/gethtml.pl?tl-QVat2-QV-BT011-028.html  
643-101198&t4-1)  
Seq primer: puc 18 forward.  
Location/Qualifiers  
source  
1..355  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT011"  
/sex="Female"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site.1: Sma1; Site.2:  
Sma1; A mini-library was made by cloning products derived  
from ORESPES PCR (U.S. Patents Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

Query Match 13.7%; Score 21; DB 9; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 111 TAAAGACCTTCAGAGAAAGA 131  
|||||  
DB 116 TAAAGACCTTCAGAGAAAGA 136

RESULT 5  
LOCUS AA234796 438 bp mRNA linear EST 06-AUG-1997  
DEFINITION z578c05.r1 Soares\_NhMMPu\_S1 Homo sapiens cDNA clone IMAGE:665512  
5', mRNA sequence.  
ACCESSION AA234796  
VERSION AA234796.1 GI:1859289  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,  
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie

TITLE 'T., Waterston,R. and Wilson,R.  
JOURNAL WashU-Merck EST Project 1997  
COMMENT Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 875 Std Error: 0.00  
Seq primer: -28m3 rev2 ET from Amersham  
High quality sequence stop: 407.  
Location/Qualifiers  
source  
1..438  
/organism="Homo sapiens"  
/db\_xref="GDB:3563479"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:669512"  
/clone\_lib="Soares\_NhMMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pTT73D-Pac  
(Pharmacia) with a modified polylinker; Site.1: Not I;  
Site.2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NDHM, pregnant uterus  
NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools of  
340488-345479, and 484488-489479."

Query Match 13.7%; Score 21; DB 9; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 111 TAAAGACCTTCAGAGAAAGA 131  
|||||  
DB 154 TAAAGACCTTCAGAGAAAGA 174

RESULT 6  
LOCUS AA456385 461 bp mRNA linear EST 06-JUN-1997  
DEFINITION aa14f05.r1 Soares\_NhMMPu\_S1 Homo sapiens cDNA clone IMAGE:813249  
5', mRNA sequence.  
ACCESSION AA456385  
VERSION AA456385.1 GI:2178961  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 461)  
AUTHORS Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,  
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie  
,T., Waterston,R. and Wilson,R.  
Contact: Wilson RK  
Unpublished (1997)  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28m3 rev2 ET from Amerisham  
 High quality sequence stop: 446.

FEATURES  
 source  
 Location/Qualifiers

1..461  
 /organism="Homo sapiens"  
 /db\_xref="GDB:6044159"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:813249"  
 /clone\_lib="Soares\_NHMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: PT733D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI. Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NDH, pregnant uterus NBHP, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 132 a 85 c 103 g 141 t

ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 Db 155 TAAAGACCTTGCAAGAAAGA 175

RESULT 7  
 BG666575 474 bp mRNA linear EST 30-APR-2001  
 LOCUS DRABNG04 Rat DRG Library Rattus norvegicus CDNA clone DRABNG04 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG666575  
 VERSION BG666575.1 GI:13888497  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 474)  
 Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,  
 Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and  
 Zhang,X.  
 Distinct gene expression profiles of rat dorsal root ganglion  
 induced by peripheral nerve axotomy  
 Unpublished (2001)  
 Contact: Zhang Xu  
 Laboratory of Sensory System  
 Institute of Neuroscience  
 320 Yue Yang Road, Shanghai 200031, P.R.China  
 Tel: 86-21-64748700-121  
 Fax: 86-21-64713446  
 Email: xu.zhang@ion.ac.cn  
 This clone is also available at Chinese National Human Genome  
 Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,  
 Pudong New Area, P.R.China. Please contact with Zhang Xu  
 (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)  
 PCR Primers  
 FORWARD: T7  
 BACKWARD: T7  
 Seq primer: T3

FEATURES  
 source  
 Location/Qualifiers

1..474  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="DRABNG04"  
 /clone\_lib="Rat DRG Library"  
 /sex="male"  
 /tissue\_type="dorsal root ganglion"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Placenta; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggccgcttcgccc); Site\_2: SfiI (ggccatttgcgc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGCGCATATGCGC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGGCGCGCATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 110 a 116 c 125 g 123 t

ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 Db 203 TAAAGACCTTGCAAGAAAGA 223

RESULT 8  
 BG573124 488 bp mRNA linear EST 10-APR-2001  
 LOCUS 602594129F1 NIH\_MGC\_79 Homo sapiens CDNA clone IMAGE:4721382 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG573124  
 VERSION BG573124.1 GI:13580777  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 488)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Clontech Laboratories, Inc.  
 CDNA Library Preparation: Clontech Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM1577 row: j column: 07  
 High quality sequence stop: 475.

FEATURES  
 source  
 Location/Qualifiers

1..488  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4721382"  
 /clone\_lib="NIH\_MGC\_79"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Placenta; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggccgcttcgccc); Site\_2: SfiI (ggccatttgcgc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGCGCATATGCGC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGGCGCGCATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 136 a 88 c 108 g 156 t

ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCGCAAGAAAGA 131  
 |||||||  
 Db 176 TAAAGACCTTCGCAAGAAAGA 196

RESULT 9  
 AM464549 489 bp mRNA linear EST 24-FEB-2000  
 LOCUS BP230016A10E9 Soares normalized bovine placenta Bos taurus CDNA  
 DEFINITION clone BP230016A10E9 5', mRNA sequence.  
 ACCESSION AM464549.1 GI:7034717  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 489)  
 Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson  
 J.H.  
 TITLE Bovine ESTs  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Lewin, H. A.  
 W. M. Keck Center for Comparative and Functional Genomics  
 University of Illinois at Urbana-Champaign  
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu  
 Funding for cattle EST sequencing was provided by the USDA National  
 Research Initiative, Animal Genome Resource Grant AG 99-3205-8534  
 to H. A. Lewin and J. E. Mowack. Base Calling/Quality Scores: PHRED  
 from Washington University Genome Center. Vector Trim: 97  
 Cross-match from Washington University Genome Center PHRAP suite.  
 Sequences submitted are vector free and at least 200 bp in length.  
 PCR Primers:  
 FORWARD: TAAAGACCTTCACCTAAGG  
 BACKWARD: ATTACCTTCACCTAAG  
 Insert Length: 489 Std Error: 0.00  
 Plate: BP230016A10 row: E column: 9  
 Seq primer: AGCGATACCAATTTCACACAGA  
 High quality sequence stop: 489.  
 Location/Qualifiers  
 1..489  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="BP230016A10E9"  
 /clone\_lib="Soares normalized bovine placenta"  
 /sex="female"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pTT3pac; Site:1; EcoRI;  
 Site:2; NotI; The cDNA library was contributed by the  
 Soares laboratory and it was constructed and normalized  
 as described by Bonaldo, M.F., Lennon, G. and Soares,  
 M.B. (1996), Genome Research 6(9): 791-806. "  
 BASE COUNT 148 a 98 c 121 g 122 t  
 ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCGCAAGAAAGA 131  
 |||||||  
 Db 403 TAAAGACCTTCGCAAGAAAGA 423

RESULT 10  
 BE198540 492 bp mRNA linear EST 26-JUN-2000  
 LOCUS u979905.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 DEFINITION IMAGE:1548632 5', mRNA sequence.  
 ACCESSION BE198540  
 VERSION BE198540.1 GI:8710709  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclerognathia; Muridae; Murinae; Mus.  
 NC1-CCAP http://www.ncbi.nlm.nih.gov/ncicgap  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:953980  
 Seq primer: -40RP from Gdbco  
 High quality sequence stop: 475.  
 Location/Qualifiers  
 1..492  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1548632"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pTT3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pTT3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 104 c 124 g 136 t  
 ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCGCAAGAAAGA 131  
 |||||||  
 Db 85 TAAAGACCTTCGCAAGAAAGA 105

RESULT 11  
 AM579514 515 bp mRNA linear EST 16-MAR-2000  
 LOCUS MR0-HR0164-070100-013-F05 HR0164 Homo sapiens CDNA, mRNA sequence.  
 DEFINITION AM579514  
 ACCESSION AM579514  
 VERSION AM579514.1 GI:7254563  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 HCCP http://www.ludwig.org.br/ORESTES.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR0&t2=MR0-HT0164-070100-013-f05&t3=2000-01-07&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 7

High quality sequence stop: 515.

#### FEATURES

##### source

Location/Qualifiers

1..515

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0164"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

low stringency conditions."

BASE COUNT 157 a 103 c 124 g 131 t

##### ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 515;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131

Db 393 TAAAGACCTTGCAAGAAAGA 413

#### RESULT 12

##### LOCUS

BM228157 560 bp mRNA linear EST 31-JAN-2002

##### DEFINITION

K0256B07-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus

##### ACCESSION

BM228157

##### VERSION

BM228157.1

##### KEYWORDS

EST.

##### SOURCE

house mouse.

##### ORGANISM

Mus musculus

##### REFERENCE

1 (bases 1 to 560)

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Stagg, C.A.,

Martin, P., Alpa, K., Tanaka, T. and Ko, M.S.H.

Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library

(Long)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

Unpublished (2001)

)"

/tissue\_type="Unfertilized Egg"

/lab\_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (http://19sun.gsc.nia.nih.gov/cDNA). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11541991]. Total RNAs were

extracted from a pool of 1488 unfertilized eggs.

Double-stranded cDNAs were synthesized with an oligo(dt)

primer (Invitrogen):

5'-GCAGTACTTCTAGATCGCGAGCGCCGCTTTT-TTTT-TTTT-3',

treated with T4 DNA polymerase, and purified by

ethanol-precipitation. The cDNAs were ligated to

lone-linker L1-SalI, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Tag polymerase (Takara) with a primer SalI-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pSPORT1 plasmid vector.

The DH10B E. coli host was transformed with the ligation

mixture by the standard chemical method. The average

insert size is about 2.5 kb. The library was constructed

by Yulan Piao (NIA)."

BASE COUNT 173 a 132 c 115 g 140 t

##### ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 560;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131

Db 480 TAAAGACCTTGCAAGAAAGA 460

#### RESULT 13

##### LOCUS

AI412929 569 bp mRNA linear EST 09-FEB-1999

##### DEFINITION

EST241229 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone

##### ACCESSION

AI412929

##### VERSION

AI412929.1

##### KEYWORDS

EST.

##### SOURCE

Rattus sp.

##### ORGANISM

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 569)

Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J.,

Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat

Gene Index

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

Unpublished (1998)

#### FEATURES

##### source

Location/Qualifiers

1..569

/organism="Rattus sp."

/db\_xref="taxon:10118"

/clone\_lib="RKEG55"

/clone\_lib="Normalized rat kidney, Bento Soares"

/note="Organ: kidney; Vector: pRT3pac; Site\_1: EcoRI;

Site\_2: NotI"

BASE COUNT 200 a 140 c 105 g 144 t  
ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||  
Db 497 TAAAGACCTTGCAAGAAAAGA 477

RESULT 14  
AM993220 594 bp mRNA linear EST 05-JUN-2000  
LOCUS RC-BN0033-180200-014-f04 BN0033 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AM993220  
VERSION AM993220.1 GI:8253368  
KEYWORDS EST  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 594)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC2-BN0033-180  
200-014-f04&ts=2000-02-18&td=1)  
Seq primer: puc 18 forward  
High quality sequence start: 20  
High quality sequence stop: 593.  
Location/Qualifiers  
1..594  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0033"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 179 a 116 c 135 g 164 t  
ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 594;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||

Db 417 TAAAGACCTTGCAAGAAAAGA 437

RESULT 15  
BF169086 597 bp mRNA linear EST 30-OCT-2000  
LOCUS 60177336AF1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3994215 5',  
DEFINITION mRNA sequence.  
ACCESSION BF169086  
VERSION BF169086.1 GI:11049438  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
1 (bases 1 to 597)  
NIH-MGC http://imgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapds@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9211 row: 0 column: 16  
High quality sequence stop: 585.  
Location/Qualifiers  
1..597  
/organism="Mus musculus"  
/strain="C57BL/6J (f1er1)"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:3994215"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
stem cell origin."  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

FEATURES  
source  
1..597  
Location/Qualifiers

BASE COUNT 175 a 106 c 143 g 173 t  
ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 597;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||  
Db 45 TAAAGACCTTGCAAGAAAAGA 65

Search completed: August 2, 2002, 22:41:40  
Job time: 30155 sec

---







alignment\_scores:  
 Quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-6 x AF006513 ..  
 Align seg 1/1 to: AF006513 from: 1 to: 5947.

1 lleleuproaspaspProaspplysysProglinalalysGlnleuglnh 17  
 |||  
 4025 ATTCTTCAGATGATCCCGATATAAAACCAACAGCAAAACAGTTGCAGAC 4074  
 17 rArgAlaAspTyrLeuIleLysLeuSerArGAspLeuAlaLys 32  
 |||  
 4075 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 4120

seq\_name: gb-pr:AC092372  
 seq\_documentation\_block:  
 LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
 DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
 ACCESSION AC092372  
 VERSION AC092372.3 GI:17402768  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 101220)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission  
 TITLE 2 (bases 1 to 101220)  
 REFERENCE DOE Joint Genome Institute.  
 AUTHORS Direct Submission  
 TITLE Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
 JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 3 (bases 1 to 101220)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 AUTHORS Direct Submission  
 TITLE Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
 JOURNAL Drive, Walnut Creek, CA 94598, USA  
 On Dec 7, 2001 this sequence version replaced gi:15290448.  
 COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 100% of Sequence;  
 Estimated Total Number of Errors is 0.  
 NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 146.7kb). It is clipped at the overlap with AC012624.  
 The number of bases overlapped is 90404.  
 FEATURES  
 source  
 1..101220  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-58M12"

BASE COUNT 34122 a 18862 c 17827 g 30409 t  
 ORIGIN

alignment\_scores:  
 Quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AC092372/rev ..

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 lleleuproaspaspProaspplysysProglinalalysGlnleuglnh 17  
 |||  
 26932 ATTCTTCAGATGATCCCGATATAAAACCAACAGCAAAACAGTTGCAGAC 26883  
 17 rArgAlaAspTyrLeuIleLysLeuSerArGAspLeuAlaLys 32  
 |||  
 26882 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 26837

seq\_name: gb-pr:AC012624  
 seq\_documentation\_block:  
 LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.  
 ACCESSION AC012624  
 VERSION AC012624.6 GI:14993679  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 134365)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission  
 TITLE 2 (bases 1 to 134365)  
 REFERENCE DOE Joint Genome Institute.  
 AUTHORS Direct Submission  
 TITLE Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint  
 JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 3 (bases 1 to 134365)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 AUTHORS Direct Submission  
 TITLE Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
 JOURNAL Drive, Walnut Creek, CA 94598, USA  
 4 (bases 1 to 134365)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 AUTHORS Direct Submission  
 TITLE Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
 JOURNAL Drive, Walnut Creek, CA 94598, USA  
 On Jul 21, 2001 this sequence version replaced gi:14277267.  
 COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1..134365  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2082117"

BASE COUNT 40414 a 24497 c 25503 g 43951 t  
 ORIGIN

alignment\_scores:  
 Quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-6 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 lleleuproaspaspProaspplysysProglinalalysGlnleuglnh 17  
 |||  
 118247 ATTCTTCAGATGATCCCGATATAAAACCAACAGCAAAACAGTTGCAGAC 118296  
 17 rArgAlaAspTyrLeuIleLysLeuSerArGAspLeuAlaLys 32  
 |||  
 118297 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 118342  
 seq\_name: gb-htg:AC021449  
 seq\_documentation\_block:

LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000  
 DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
 pieces.  
 AC021449  
 AC021449.3 GI:10047806  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 143079)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens, clone RP11-58M12  
 Unpublished  
 2 (bases 1 to 143079)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Chepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
 Deatellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPeeters,R., Meldrim,J., Menes,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,  
 Stojanovic,N., Sudramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:1407963.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5154  
 Center clone name: 58\_M12  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 134743 bases at least Q40  
 Consensus quality: 139227 bases at least Q30  
 Consensus quality: 140814 bases at least Q20  
 Insert size: 144000; agarose-1p  
 Insert size: 142179; sum-of-ctrls  
 Quality coverage: 4.6 in Q20 bases; agarose-1p  
 Quality coverage: 4.7 in Q20 bases; sum-of-ctrls  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 38820: contig of 38820 bp in length  
 \* 38921 38920: gap of 100 bp  
 \* 38921 40411: contig of 1491 bp in length  
 \* 40412 40511: gap of 100 bp  
 \* 40512 43279: contig of 2768 bp in length  
 \* 43280 43379: gap of 100 bp

FEATURES  
 source  
 1. 143079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-58M12"  
 /clone\_1b="RP11-58M12 Human Male BAC"  
 1. 38820  
 /note="assembly-fragment  
 clone\_end:SP6  
 vector\_side:left"  
 38921..40411  
 /note="assembly-fragment"  
 40512..43279  
 /note="assembly-fragment"  
 43380..46905  
 /note="assembly-fragment"  
 47006..51830  
 /note="assembly-fragment"  
 51931..62619  
 /note="assembly-fragment"  
 62720..75408  
 /note="assembly-fragment"  
 75509..92516  
 /note="assembly-fragment"  
 92617..106409  
 /note="assembly-fragment"  
 106510..143079  
 /note="assembly-fragment  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
 ORIGIN  
 alignment\_scores:  
 quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AC021449 ..  
 Align seg 1/1 to: AC021449 from: 1 to: 143079  
 1 IileuproaspaspProaspplvysProGlnAlalySglnLeuGlnth 17  
 |||||||  
 116147 ATCTTCCAGATGATCCCGATTAACACACACACACACACACAC 116196  
 |||||||  
 17 TARGALASPTyRleuileuLeuLeuSerArgAspLeuAlaLys 32  
 |||||||  
 116197 CCGGACAGACTACCTCATCAATTAAGTAGAGAGACTTGCACAA 116242  
 |||||||  
 seq\_name: gb\_htg:AC008531  
 seq\_documentation\_block:  
 LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
 7 ordered pieces.  
 AC008531  
 AC008531.3 GI:12830078



```
* 12561 14611: contig of 2051 bp in length
* 14612 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 16481: gap of unknown length
* 16482 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
* 20435 20535: gap of unknown length
* 20536 23515: contig of 2981 bp in length
* 23516 23616: gap of unknown length
* 23617 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 67981: gap of unknown length
* 67982 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87612: gap of unknown length
* 87613 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
    source          1..193446
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="RP11-36012"
                    /clone_lib="RP11 human BAC library 11"
BASE COUNT      57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
alignment_scores:
    Quality:      32.00      Length:      32
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
    US-08-973-363-6 x AC091946 ..
Align seg 1/1 to: AC091946 from: 1 to: 193446
1 11leuProaSpProaSpLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
87658 ATTCTTCAGATGATCCGATATAAAACACACAAACAGTTCGACAGC 87707
```

```
17 fARGAlaSpTyrLeuIleLysLeuSerArgAspLeuAlaLys 32
|||||
87708 CCGTGACAGACTACCTCATCAATTTACTTAGAGAGATCTTGCAAAA 87753
seq_name: gb_pr:AC026778
seq_documentation_block:
    LOCUS      AC026778      195433 bp      DNA      linear      PRI 01-JUN-2001
    DEFINITION  Homo sapiens chromosome 5 clone CTC-428111, complete sequence.
    ACCESSION  AC026778
    VERSION    AC026778.4 GI:14277282
    KEYWORDS   HTG.
    SOURCE      human.
    ORGANISM   Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 195433)
                DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS       DOE Joint Genome Institute
TITLE         Unpublished
JOURNAL       2 (bases 1 to 195433)
REFERENCE     2 (bases 1 to 195433)
                DOE Joint Genome Institute.
TITLE         Direct Submission
JOURNAL       Direct Submission
AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL       Drive, Walnut Creek, CA 94598, USA
AUTHORS       On Jun 1, 2001 this sequence version replaced gi:13677045.
TITLE         Draft Sequence Produced by DOE Joint Genome Institute
JOURNAL       www.jgi.doe.gov
COMMENT       Finishing Completed at Stanford Human Genome Center
                www.shgc.stanford.edu
                Quality: Phrap Quality >=40 99.9% of Sequence;
                Estimated Total Number of Errors is 0.2.
                SFS Content:
                WI-13675 G23101
                SHGC-58345 G38487
                SHGC-103595 G57841.
FEATURES
    source          1..195433
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTC-428111"
BASE COUNT      62762 a 37302 c 37040 g 56329 t
ORIGIN
alignment_scores:
    Quality:      32.00      Length:      32
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
    US-08-973-363-6 x AC026778/rev ..
Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433
1 11leuProaSpProaSpLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
29732 ATTCTTCAGATGATCCGATATAAAACACACAAACAGTTCGACAGC 29683
17 fARGAlaSpTyrLeuIleLysLeuSerArgAspLeuAlaLys 32
|||||
29682 CCGTGACAGACTACCTCATCAATTTACTTAGAGAGATCTTGCAAAA 29637
seq_name: gb_pr:AC022121
seq_documentation_block:
```

```

LOCUS       AC022121      219258 bp    DNA             linear    PRI 30-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION   AC022121
VERSION     AC022121.6   GI:15375145
KEYWORDS    SOURCE
ORGANISM    human
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE   1 (bases 1 to 219258)
AUTHORS     Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
TITLE       DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL     Direct Submission
REFERENCE   2 (bases 1 to 219258)
AUTHORS     Unpublished
TITLE       DOE Joint Genome Institute.
JOURNAL     Direct Submission
REFERENCE   Submitted (26-JUN-2000) Production Sequencing Facility, DOE Joint
AUTHORS     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE       3 (bases 1 to 219258)
JOURNAL     DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE   Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS     Drive, Walnut Creek, CA 94598, USA
TITLE       4 (bases 1 to 219258)
JOURNAL     DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE   Direct Submission
AUTHORS     Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
TITLE       Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 30, 2001 this sequence version replaced g1:15148108.
             Draft sequence produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing completed at Stanford Human Genome Center
             www.shgc.stanford.edu
             Quality: Phrap Quality >=40.99.8% of sequence;
             Estimated Total Number of Errors is 0.4.
             STS Content:
             STS 5811.G04.974
             WI-13675.G23101
             SHGC-58345.G38487
             SHGC-103595.G57841.
FEATURES             Location/Qualifiers
     source          1..219258
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2007H13"
BASE COUNT       71954 a 42062 c 40933 g 64309 t
ORIGIN
alignment_scores:
    Quality:      32.00      Length:      32
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000  Percent Identity: 100.000
Alignment block:
US-08-973-363-6 x AC022121/rev ..
Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258
1  ||leleuproaspapspProaspLysLysProglAlaLysGlnleuGlnp 17
   |||||||
89450 ATTCTTCACGATGATCCCATTAACCAACCAACCAACCACTTCGCAGC 89401
17 rAgaLaapPyrlleuLleLysleuLysSerArgaLpLeuAlaLys 32
   |||||||
89400 CCGGCGACACTCCCTCATCAATTACTTAGTAGAGATCTTGCAAA 89355
seq_name: gb:htg:AC092382
seq_documentation_block:
LOCUS       AC092382      276181 bp    DNA             linear    HNG 03-JUL-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE.
ACCESSION   AC092382
VERSION     HTG: HTGSEQ1: HTG_SEQ1: HTG_ACTIVEFIN.
KEYWORDS    SOURCE
ORGANISM    human
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE   1 (bases 1 to 276181)
AUTHORS     Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
TITLE       DOE Joint Genome Institute.
JOURNAL     Sequencing of Human Chromosome 5
REFERENCE   2 (bases 1 to 276181)
AUTHORS     Unpublished
TITLE       DOE Joint Genome Institute.
JOURNAL     Direct Submission
REFERENCE   Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
AUTHORS     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE       -----Genome Center
JOURNAL     Center: Joint Genome Institute
COMMENT     Center Code: JGI
             Web site: http://www.jgi.doe.gov
             -----
             Project Information
             Center Project Name: 435334
             Center clone name: RP11-75H1
             -----
             Summary Statistics
             Consensus quality: 229677 bases at least Q40
             Consensus quality: 256163 bases at least Q30
             Consensus quality: 260799 bases at least Q20
             Estimated insert size: 174820; agarose-fp estimation
             Estimated insert size: 271581; sum-of-contigs estimation
             Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
             Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.
             NOTE: This is a 'working draft' sequence. It currently
             * consists of 47 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             1
             1061: contig of 1061 bp in length
             1062
             1161: gap of unknown length
             1162
             2827: contig of 1666 bp in length
             2828
             2927: gap of unknown length
             2928
             4227: contig of 1300 bp in length
             4228
             4327: gap of unknown length
             4328
             5618: contig of 1291 bp in length
             5619
             5718: gap of unknown length
             5719
             6983: contig of 1265 bp in length
             6984
             7083: gap of unknown length
             7084
             8422: contig of 1339 bp in length
             8423
             8523: gap of unknown length
             8524
             9771: contig of 1249 bp in length
             9772
             9871: gap of unknown length
             9872
             11198: contig of 1327 bp in length
             11199
             11298: gap of unknown length
             11299
             12414: contig of 1116 bp in length
             12415
             12514: gap of unknown length
             12515
             13806: contig of 1292 bp in length
             13807
             13906: gap of unknown length
             13907
             15360: contig of 1454 bp in length
             15361
             15460: gap of unknown length
             15461
             16687: contig of 1207 bp in length
             16688
             16767: gap of unknown length
             16768
             18390: contig of 1623 bp in length
             18391
             18490: gap of unknown length
             18491
             19541: contig of 1051 bp in length
             19542
             19641: gap of unknown length
             21156: contig of 1515 bp in length
             21157
             21256: gap of unknown length
             22448: contig of 1192 bp in length

```

\* 22449 22548: gap of unknown length  
\* 22549 24499: contig of 1951 bp in length  
\* 24500 24599: gap of unknown length  
\* 24600 26384: contig of 1785 bp in length  
\* 26385 26484: gap of unknown length  
\* 26485 27883: contig of 1399 bp in length  
\* 27884 27983: gap of unknown length  
\* 27984 30386: contig of 2403 bp in length  
\* 30387 30486: gap of unknown length  
\* 30487 32480: contig of 1994 bp in length  
\* 32481 32580: gap of unknown length  
\* 32581 35441: contig of 2861 bp in length  
\* 35442 35541: gap of unknown length  
\* 35542 37800: contig of 2259 bp in length  
\* 37801 37900: gap of unknown length  
\* 37901 40587: contig of 2687 bp in length  
\* 40588 40687: gap of unknown length  
\* 40688 44159: contig of 3472 bp in length  
\* 44160 44259: gap of unknown length  
\* 44260 46366: contig of 2377 bp in length  
\* 46367 46736: gap of unknown length  
\* 46737 50082: contig of 3346 bp in length  
\* 50083 50182: gap of unknown length  
\* 50183 53988: contig of 3806 bp in length  
\* 53989 54088: gap of unknown length  
\* 54089 56592: contig of 2504 bp in length  
\* 56593 56692: gap of unknown length  
\* 56693 61352: contig of 4660 bp in length  
\* 61353 61452: gap of unknown length  
\* 61453 64254: contig of 2802 bp in length  
\* 64255 64354: gap of unknown length  
\* 64355 68825: contig of 4471 bp in length  
\* 68826 68925: gap of unknown length  
\* 68926 74393: contig of 5468 bp in length  
\* 74394 74493: gap of unknown length  
\* 74494 80561: contig of 6068 bp in length  
\* 80562 80661: gap of unknown length  
\* 80662 87626: contig of 6965 bp in length  
\* 87627 87726: gap of unknown length  
\* 87727 93600: contig of 5874 bp in length  
\* 93601 93700: gap of unknown length  
\* 93701 102024: contig of 8324 bp in length  
\* 102025 102124: gap of unknown length  
\* 102125 109905: contig of 7781 bp in length  
\* 109906 110005: gap of unknown length  
\* 110006 116737: contig of 6732 bp in length  
\* 116738 116837: gap of unknown length  
\* 116838 126797: contig of 9960 bp in length  
\* 126798 126897: gap of unknown length  
\* 126898 135114: contig of 8217 bp in length  
\* 135115 135214: gap of unknown length  
\* 135215 146713: contig of 11499 bp in length  
\* 146714 146813: gap of unknown length  
\* 146814 157894: contig of 11081 bp in length  
\* 157895 157994: gap of unknown length  
\* 157995 171752: contig of 13758 bp in length  
\* 171753 171852: gap of unknown length  
\* 171853 201687: contig of 29835 bp in length  
\* 201688 201787: gap of unknown length  
\* 201788 231268: contig of 29481 bp in length  
\* 231269 231368: gap of unknown length  
\* 231369 276181: contig of 44813 bp in length.

FEATURES  
source

1..276181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-75H1"  
/clone.lib="RPci human BAC library 11"  
BASE COUNT 86859 a 51769 c 50026 g 82854 t 4673 others  
ORIGIN

alignment\_scores:  
Quality: 32.00 Length: 32  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x AC092382/rev ..

Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181

1 TleleuProaspaspProaspLysProglAlaLysGlnLeuGlnH 17  
|||||  
213694 ATTCTCCAGATGATCCGATFAAAAAACCAACCAACAGATTGCAGAC 213645

17 rArgAlaAspTyrlleuLleLysLeuLeuSerArgAspLeuAlaLys 32  
|||||  
213644 CCGTGCAGACTACTCTCATTAATTACTTAGTAGAGATTGCAGAAA 213599

seq\_name: gb\_pat:A58696

seq\_documentation\_block:

LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 15 from Patent W09639505.

ACCESSION A58696  
VERSION A58696.1 GI:3714253

KEYWORDS  
SOURCE

ORGANISM  
unidentified.  
unclassified.

REFERENCE  
1 (bases 1 to 1316)

AUTHORS  
Griffiths,R. and Tiwari,B.

TITLE  
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS

JOURNAL  
Patent: WO 9639505-A 15 12-DEC-1996;

COMMENT  
ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.

FEATURES  
source 1..1316  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x A58696 ..

Align seg 1/1 to: A58696 from: 1 to: 1316

1 TleleuProaspaspProaspLysProglAlaLysGlnLeuGlnH 17  
|||||  
1138 ATTCTCCAGATGATCCGATFAAAAAACCAACCAACAGATTGCAGAC 1187

17 rArgAlaAspTyrlleuLleLysLeuLeuSerArgAspLeuAlaLys 26  
|||||  
1188 TCGTGCAGATTACTCTCATTAATTACTG 1215

seq\_name: gb\_pat:A58691

seq\_documentation\_block:

LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent W09639505.

ACCESSION A58691  
VERSION A58691.1 GI:3714250

KEYWORDS  
SOURCE

ORGANISM  
unidentified.  
unclassified.

REFERENCE	1 (bases 1 to 6608)
AUTHORS	Griffiths R. and Tjweri, B.
TITLE	AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL	Patent: WO 9639505-A 10 12-DEC-1996;
COMMENT	ISIS INNOVATION (GB)
FEATURES	Other publication NO 5966996 961224.
SOURCE	Location/Qualifiers 1..6608
BASE COUNT	2289 a 1207 c 1459 g 1653 t
ORIGIN	
Alignment_scores:	
Quality:	26.00 Length: 26
Ratio:	1.000 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 100.000
Alignment_block:	
US-08-973-363-6 x A58691	..
Align seq 1/1 to: A58691 from: 1 to: 6608	
1	11leuProaaspAproaspLystysProGlnalAlatysInleuGln17
4080	ATTTTACCGATGATCCAGACAGAAACCCAGGCAAGCCTACAGAC 4129
17	TATGAlaAAPTyrleu1leYleu1leu 26
4130	CCGTCGACACTACCTCATTAATTATCTGC 4157
seq_name: gb_ov:AF004397	
seq_documentation_block:	
LOCUS	AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION	Gallus gallus chomo-helicase-DNA-binding on the Z chromosome
DESCRIPTION	protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
ACCESSION	AF004397
VERSION	AF004397.1 GI:2501845
KEYWORDS	chd.
SOURCE	chicken.
ORGANISM	Gallus gallus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
	Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 6872)
AUTHORS	Griffiths R. and Korn, R.M.
TITLE	A CHD1 gene is 2 chromosome linked in the chicken Gallus domesticus
JOURNAL	Gene 197 (1-2), 225-229 (1997)
MEDLINE	97473516
REFERENCE	2 (bases 1 to 6872)
AUTHORS	Griffiths R. and Korn, R.M.
TITLE	Direct Submission
JOURNAL	Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University
FEATURES	Location/Qualifiers 1..6872
SOURCE	1..6872
gene	/organism="Gallus gallus"
	/db_xref="taxon:9031"
	1..6872
CDS	/gene="CHD-Z"
	228..5654
	/function="role in chromatin architecture"
	/note="CHD protein with hydrophilic domain"
	/codon_start=1
	/product="chomo-helicase-DNA-binding on the Z chromosome
	protein
	/protein_id="AAC60282.1"
	/db_xref="GI:2501846"

[illegible]

```

source 1.1311
BASE COUNT 441 a 257 c 316 g 297 t
ORIGIN

alignment_scores:
  Quality: 25.00 Length: 25
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AR029026 ..

Align seg 1/1 to: AR029026 from: 1 to: 1311

1 11leuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnH 17
|||||
1237 ATTCTTCAGATGATCCTGATATAAACCACAGACAAACAGTTACAGAC 1286
|||||
17 rArgAlaaspTYrLeu1leLysLeu 25
|||||
1287 CCCTGCAGACTACCTCATCAACTA 1311

```





8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeu 24  
 |||  
 52 AAAAAACCAAGCAAAACAGTTACAGACCCGTCAGACTACTCTATCAA 101  
 |||  
 24 sLeuLeuSerArgAspLeuAlaLysArgGlnAlaGlnArgLeuCysGlyA 41  
 |||  
 102 ACTACTTACGACAGATCTTGCAGAAAAGAGAGCGCTCAGAGACTTTGTGTG 151  
 |||  
 41 la 41  
 ||  
 152 CG 153

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142754

seq\_documentation\_block:

ID AA142754 standard; cDNA: 1316 BP.

AC AA142754;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-W gene (partial sequence).

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN WO9639505-A1.

PD 12-DEC-1996.

PE 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny

PS Claim 1; Fig 8; 76pp: English.

CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-W was deduced  
 CC from 2 clones isolated from a 10-day chick embryo library using  
 CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is  
 CC located on the W chromosome. Probes based on CHD-W and CHD-1A give  
 CC a W chromosome-specific signal on hybridisation to genomic DNA of a  
 CC non-rafite bird and can be used for sex determin. of a bird. CHD-W  
 CC nucleic acids can also be used to control the sex of progeny of a  
 CC bird.

SO Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

alignment\_scores:

Quality: 26.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AA142754 ..

Align seg 1/1 to: AA142754 from: 1 to: 1316

1 lIleuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17

|||||  
 1138 ATTTCCTGATGATCCAGTAAAGAAACCCAGGCTAGAGCTTACAGAC 1187  
 |||  
 17 fArgAlaAspTyrLeuIleLysLeuLeu 26  
 |||  
 1188 TCCTGCAGATTACTCTATTAAATTACTG 1215

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA: 6608 BP.

AC AA142751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN WO9639505-A1.

PD 12-DEC-1996.

PE 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny

PS Claim 1; Fig 5; 76pp: English.

CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
 CC library using a great tit CHD-W sequence (see also AA142755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AA142756-57). It is located on an autosome or X  
 CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
 CC specific signal on hybridisation to genomic DNA of a non-rafite  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
 CC acids can also be used to control the sex of the progeny of a bird.

SO Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 26.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lIleuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
 |||

4080 ATTTCAGTATGATCCAGACAAAGAACCCAGGCAAGACGCTACAGAC 4129

17 rArgalaaspTyrlleuileysleu 26

4130 CCGTCGACGACTCCTCATTAATACAG 4157

seq\_name: /SIDSL/gcgdata/hold-geneseq/genesegn-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; CDNA; 1311 BP.

AAV59280;

14-DEC-1998 (first entry)

Altered telomere repeat binding factor 1 gene.

ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;

telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

Home sapiens.

Synthetic.

Key

CDS

1.1311

/tag- a

/product- "A-TRF"

MO9836066-A1.

20-AUG-1998.

13-FEB-1998; 98MO-US02765.

04-FEB-1998; 9805-0018628.

13-FEB-1997; 9705-0800264.

(UVRQ ) UNIV ROCKEFELLER.

Blanchi A, De Lange T, Van Steensel B;

WPI; 1998-480769/41.

P-PSDB; AAM59280.

Nucleic acid encoding altered telomere repeat binding protein and

related vectors - transformants, hetero-dimers and antibodies, used

to inhibit shortening of telomerases caused by ageing or disease,

also used to extend life of cells in culture

Claim 14; Page 110-111; 163pp; English.

The altered vertebrate telomere repeat binding protein (A-TRF) has a

telomere repeat binding factor (TRF) dimerisation domain, and forms a

hetero-dimer with TRF, preventing it from binding to the specified repeat

sequence. A-TRF, optionally expressed by gene therapy, is used to

inhibit shortening of telomeres associated with ageing (for cosmetic

purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,

atrophy of the skin, age-related macular degeneration, atherosclerosis,

tumours and viral (including human immune deficiency virus) infection.

Cells expressing A-TRF also have an increased life span in vitro, e.g.

for expression of recombinant proteins or where intended for subsequent

transplant or for testing, eliminating the need for transformation.

Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

Alignment-scores:

Quality: 25.00 Length: 25

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

Alignment-block:

US-08-973-363-6 x AAV59280 ..

Align seg 1/1 to: AAV59280 from: 1 to: 1311

1 lIleuPProAspPProAspLysLysProGlnAlaGlnleuGlnth 17

1237 ATTCTTCAGATGATCCTCATTAATAAACCCAGGCAAGACGCTACAGAC 1286

17 rArgalaaspTyrlleuileysleu 25

1287 CCGTCGACGACTCCTCATTAATACAG 1311

seq\_name: /SIDSL/gcgdata/hold-geneseq/genesegn-emb1/NA1997.DAT:AAAT42759

seq\_documentation\_block:

ID AAT42759 standard; DNA; 153 BP.

AAAT42759;

12-MAR-1997 (first entry)

Great tit CHD-W gene fragment.

Bird; sex determination; chromodomain-Helicase-DNA binding 1;

CHD-1A; CHD-W; W chromosome; ss.

Parus major.

Key

misc\_difference 52..81

/tag- a

/note- "bases 52-81 are a repeat of bases 22-51

and are ignored in the translated amino

acid sequence given in Fig 3"

MO9639505-A1.

12-DEC-1996.

05-JUN-1996; 96MO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Giffiths R, Tiwari B;

WPI; 1997-043127/04.

P-PSDB; AAM08149.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Claim 8; Fig 3; 76pp; English.

Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology

to portions of the chicken CHD-1A (A - Avian) gene (AAT42757),

chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and

and the great tit CHD-W gene (AAT42759). Translated amino acid

sequences of this region are provided in AAM08146-49. The CHD-1A

(see also AAT42751) and CHD-W (see also AAT42754-55) genes determine

sex in birds and can be used to identify the sex of an embryo,

foetus etc. and to manipulate the sex of progeny.

Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

Alignment-scores:

Quality: 19.00 Length: 19

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

Alignment-block:

US-08-973-363-6 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
|||||
52 AAGAAACCCAGGCAAGACAGTGTGACAGCCGTCAGATTCATCTATTA 101
24 sleuLeu 26
|||||
102 ATTACTG 108
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42757

seq\_documentation\_block:

ID AAT42757 standard; DNA; 153 BP.

AC AAT42757;

DT 12-MAR-1997 (first entry)

DE Chick CHD-1A gene fragment.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag- a

FT /note- "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96MO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAW08147.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

PS Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757).  
CC chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758)  
CC and the great tit CHD-W gene (see also AAT42759). Translated amino  
CC acid sequences of this region are provided in AAW08146-49. The  
CC CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes  
CC determine sex in birds and can be used to identify the sex of an  
CC embryo, foetus etc. and to manipulate the sex of progeny.

CC Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 19.00

Length: 19

Ratio: 1.000

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAT42757 ..

Align seg 1/1 to: AAT42757 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
|||||
52 AAGAAACCCAGGCAAGACAGTGTGACAGCCGTCAGATTCATCTATTA 101
24 sleuLeu 26
|||||
102 ATTACTG 108
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42758

seq\_documentation\_block:

ID AAT42758 standard; DNA; 153 BP.

AC AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag- a

FT /note- "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96MO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAW08148.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

PS Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757).  
CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
CC sequences of this region are provided in AAW08146-49. The CHD-1A  
CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
CC sex in birds and can be used to identify the sex of an embryo,  
CC foetus etc. and to manipulate the sex of progeny.

CC Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment\_scores:

Quality: 19.00

Length: 19

Ratio: 1.000

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAT42758 ..

```

Align seg 1/1 to: AAT42758 from: 1 to: 153
      8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24
      |||
      52 AAGAAACCCCGAGCTAAGCAGTTCAGACCCGTCGAGATTCCTCATTTAA 101
      24 sLeuLeu 26
      |||
      102 ATTACTG 108

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL06443
seq_documentation_block:
ID   ABL06443 standard; cDNA; 6240 BP.
XX
AC   ABL06443;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
XX   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
XX   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
DR   WPI; 2001-656860/75.
XX   P-PSDB; ABB62340.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
XX   genes from Drosophila and for elucidating cell signalling and cell-cell
XX   interactions -
XX
PS   Claim 1; SEQ ID NO 13811; 21bp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
XX   capable of detecting 1000 or more genes from Drosophila. The invention is
XX   useful in developmental biology and in elucidating cell signalling and
XX   cell-cell interactions in higher eukaryotes for the development of
XX   insecticides, therapeutics and pharmaceutical drugs. The invention
XX   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX   sequences (ABBS7737-ABB72072).
XX   (ABBS7737-ABB72072).
XX
CC   The sequence data for this patent did not form part of the printed
XX   specification, but was obtained in electronic format directly from WIPO
XX   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

alignment_scores:
      Quality: 11.00      Length: 11
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```

```

      9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19
      |||
      4205 AAGCCCCAGCCCAAGCAGCTGCAGACGGCTGCC 4237

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL06442
seq_documentation_block:
ID   ABL06442 standard; cDNA; 9933 BP.
XX
AC   ABL06442;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
XX   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
XX   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
DR   WPI; 2001-656860/75.
XX   P-PSDB; ABB62339.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
XX   genes from Drosophila and for elucidating cell signalling and cell-cell
XX   interactions -
XX
PS   Claim 1; SEQ ID NO 13808; 21bp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
XX   capable of detecting 1000 or more genes from Drosophila. The invention is
XX   useful in developmental biology and in elucidating cell signalling and
XX   cell-cell interactions in higher eukaryotes for the development of
XX   insecticides, therapeutics and pharmaceutical drugs. The invention
XX   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX   sequences (ABBS7737-ABB72072).
XX   (ABBS7737-ABB72072).
XX
CC   The sequence data for this patent did not form part of the printed
XX   specification, but was obtained in electronic format directly from WIPO
XX   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

alignment_scores:
      Quality: 11.00      Length: 11
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933
      9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19
      |||
      6745 AAGCCCCAGCCCAAGCAGCTGCAGACGGCTGCC 6777

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAK88882

```

seq\_documentation\_block:  
ID AAK8882 standard; cDNA; 421 BP.  
XX  
AC AAK8882;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 08-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.



PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239937.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0246174.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-457727/49.  
XX  
XX P-PSDB; AAM38625.  
XX  
XX  
XX Isolated polypeptide for treating, preventing and/ or prognosing  
XX  
XX PT disorders related to the colon and rectum including colorectal cancers  
XX  
XX PT and also for testing and detection e.g. diagnosis -

XX  
XX Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in  
CC AA157547-AA157619 and AAM38569-AAM38641. These can be used in the  
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
CC The present sequence is a colorectal cancer antigen coding sequence of  
CC the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment block:  
US-08-973-363-6 x AA157603 ..  
  
Align seq 1/1 to: AA157603 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
146 AAGCAGCTACGACCCGAGCGATTACTTG 175  
  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAC99822  
  
seq\_documentation\_block:  
ID AAC99822 standard; cDNA; 1148 BP.  
XX  
XX AAC99822;  
XX  
XX 13-MAR-2001 (first entry)  
XX  
XX Human secreted protein gene 5 SEQ ID NO:15.  
XX  
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
XX antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;  
XX cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
XX fungicide; ophthalmological; gene therapy; pathological condition;  
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
XX neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
XX cerebrovascular disorder; angiogenesis; nervous system disorder;  
XX Alzheimer's disease; infection; ocular disorder; corneal infection;  
XX wound healing; skin aging; food additive; preservative; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200070042-A1.  
XX  
XX 23-NOV-2000.  
XX  
XX 11-MAY-2000; 2000WO-US12788.  
XX  
XX 13-MAY-1999; 99US-0134068.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
XX  
XX Duan RD, Florence KA, Soppet DR;  
XX  
XX WPI; 2000-679628/66.  
XX  
XX P-PSDB; AAB56081.  
XX  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
XX  
XX PT used in preventing, treating or ameliorating a medical condition -  
XX  
XX PS Claim 1; Page 858; 1065pp; English.



```
XX CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC CC human secreted proteins given in AAB56077 to AAB56562. Human secreted
CC CC proteins have activities based on the tissues and cells the genes are
CC CC expressed in. Examples of activities include: immunosuppressive;
CC CC antitachytic; antithrombotic; antiproliferative; cytostatic; cardiant;
CC CC vasotrophic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC CC virucide; fungicide; and ophthalmological. The human secreted
CC CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC CC condition or susceptibility to a pathological condition. Disorders which
CC CC are diagnosed or treated include autoimmune diseases e.g. Rheumatoid
CC CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC CC be used to aid wound healing and epithelial cell proliferation, to
CC CC prevent skin aging due to sunburn, to maintain organs before
CC CC transplantation, for supporting cell culture of primary tissues, to
CC CC regenerate tissues and in chemotherapy. The proteins can also be used as a
CC CC food additive or preservative to increase or decrease storage
CC CC capabilities AAC99809 to AAC99817 and AAB56076 represent sequences used
CC CC in the exemplification of the present invention.
XX CC
SQ Sequence 1148 BP; 353 A; 251 C; 221 G; 322 T; 1 other:

alignment_scores:
      Quality: 8.00      Length: 8
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AAC99822 ..

Align seg 1/1 to: AAC99822 from: 1 to: 1148

27 SerArgAspLeuAlaIysArgGlu 34
|||||
162 TCTAGGATCTAGCAAGCGAGG 185

seq_name: /SIS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS96223

seq_documentation_block:
ID AAS96223 standard; CDNA; 1973 BP.
XX AC
XX AAS96223:
XX
XX 26-FEB-2002 (first entry)
XX
XX DE Human bone marrow tissue polynucleotide #9.
XX
XX KW Human; bone marrow; central nervous system disorder; haematopoiesis; ss;
XX KW peripheral nervous system disorder; myeloid cell disorder; osteoporosis;
XX KW lymphoid cell disorder; osteoarthritis; bone degenerative disorder; skin;
XX KW periodontal disease; cerebrovascular disorder; viral infection; cancer;
XX KW bacterial infection; fungal infection; autoimmune disorder; pancreas;
XX KW hyperproliferative disorder; cognitive depression; stomach;
XX KW inflammatory disorder; tumour; colon; liver; bladder; prostate; kidney;
XX KW ovary; brain; skin; bone; cartilage; tendon; ligament; nerve;
XX KW tissue growth; tissue regeneration; wound; burn; incision; ulcer.
XX
XX OS Homo sapiens.
XX
XX PN WO200179447-A2.
XX
XX PD 25-OCT-2001.
XX
XX PF 18-APR-2001; 2001WO-US12607.
XX
XX PR 18-APR-2000; 2000US-0522929.
```

```
PR 24-OCT-2000; 2000US-0695783.
PR 30-NOV-2000; 2000US-250583P.
PR 26-JAN-2001; 2001US-0770160.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Ford JE, Boyle BJ, Tang YT, Asundi V, Yang Y, Liu C, Drmanac RT;
XX DR WPI; 2001-626527/72.
XX DR P-PSDB; AAU71931.
XX
XX PT Polynucleotides encoding bone marrow tissue polypeptides, useful for
XX PS preventing, diagnosing and treating osteopathic diseases -
XX
XX PS Claim 1; Page 137-138; 202pp; English.
XX
XX CC The invention relates to isolated polynucleotides encoding bone marrow
XX CC tissue polypeptides. The sequences of the invention can be used in the
XX CC treatment of central nervous system disorders such as Parkinson's
XX CC disease, Alzheimer's disease, Huntington's disorders and amyotrophic
XX CC lateral sclerosis, peripheral nervous system disorders such as peripheral
XX CC neuropathy, haematopoiesis, myeloid and lymphoid cell disorders,
XX CC osteoporosis, osteoarthritis, bone degenerative disorders, periodontal
XX CC disease, cerebrovascular disorders such as stroke, viral infections,
XX CC bacterial infections, fungal infections, autoimmune disorders such as
XX CC rheumatoid arthritis, multiple sclerosis, asthma and eczema, inflammatory
XX CC disorders such as Crohn's disease and nephritis, hyperproliferative
XX CC disorders such as psoriasis, cognitive disorders, depression and cancer
XX CC (e.g. tumours of the stomach, colon, pancreas, liver, bladder, prostate,
XX CC ovary, kidney, brain and skin). The sequences are also useful in bone,
XX CC cartilage, tendon, ligament and/or nerve tissue growth or regeneration
XX CC and in healing of wounds, burns, incisions and ulcers. Sequences
XX CC AAC96215-AAS96271 represent cDNA molecules encoding bone marrow tissue
XX CC polypeptides of the invention.
XX
XX SQ Sequence 1973 BP; 604 A; 338 C; 352 G; 679 T; 0 other:

alignment_scores:
      Quality: 8.00      Length: 8
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AAS96223 ..

Align seg 1/1 to: AAS96223 from: 1 to: 1973

27 SerArgAspLeuAlaIysArgGlu 34
|||||
1683 TCTAGGATCTAGCAAGCGAGG 1706

seq_name: /SIS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA71645

seq_documentation_block:
ID AAA71645 standard; DNA; 2052 BP.
XX AC
XX AAA71645:
XX
XX DT 14-DEC-2000 (first entry)
XX
XX DE Human aspartate protease psl 5 protein encoding DNA.
XX
XX KW Aspartate protease; neuroprotectant; nootropic; gamma-secretase activity;
XX KW amyloid precursor protein; APP; beta-amyloid; cleavage; Notch; Irel;
XX KW Alzheimer's disease; psl 5; human; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200043505-A2.
XX
XX PD 27-JUL-2000.
```



27-SEP-2000: 2000US-0235834.  
 PR 27-SEP-2000: 2000US-0235836.  
 PR 29-SEP-2000: 2000US-0236327.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236368.  
 PR 29-SEP-2000: 2000US-0236369.  
 PR 29-SEP-2000: 2000US-0236370.  
 PR 02-OCT-2000: 2000US-0236802.  
 PR 02-OCT-2000: 2000US-0237037.  
 PR 02-OCT-2000: 2000US-0237038.  
 PR 02-OCT-2000: 2000US-0237039.  
 PR 02-OCT-2000: 2000US-0237040.  
 PR 13-OCT-2000: 2000US-0239935.  
 PR 13-OCT-2000: 2000US-0239937.  
 PR 20-OCT-2000: 2000US-0240960.  
 PR 20-OCT-2000: 2000US-0241221.  
 PR 20-OCT-2000: 2000US-0241785.  
 PR 20-OCT-2000: 2000US-0241786.  
 PR 20-OCT-2000: 2000US-0241787.  
 PR 20-OCT-2000: 2000US-0241808.  
 PR 20-OCT-2000: 2000US-0241809.  
 PR 01-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0246474.  
 PR 08-NOV-2000: 2000US-0246475.  
 PR 08-NOV-2000: 2000US-0246476.  
 PR 08-NOV-2000: 2000US-0246477.  
 PR 08-NOV-2000: 2000US-0246478.  
 PR 08-NOV-2000: 2000US-0246523.  
 PR 08-NOV-2000: 2000US-0246524.  
 PR 08-NOV-2000: 2000US-0246525.  
 PR 08-NOV-2000: 2000US-0246526.  
 PR 08-NOV-2000: 2000US-0246527.  
 PR 08-NOV-2000: 2000US-0246528.  
 PR 08-NOV-2000: 2000US-0246532.  
 PR 08-NOV-2000: 2000US-0246609.  
 PR 08-NOV-2000: 2000US-0246610.  
 PR 08-NOV-2000: 2000US-0246611.  
 PR 08-NOV-2000: 2000US-0246613.  
 PR 17-NOV-2000: 2000US-0248207.  
 PR 17-NOV-2000: 2000US-0248208.  
 PR 17-NOV-2000: 2000US-0248209.  
 PR 17-NOV-2000: 2000US-0249210.  
 PR 17-NOV-2000: 2000US-0249211.  
 PR 17-NOV-2000: 2000US-0249212.  
 PR 17-NOV-2000: 2000US-0249213.  
 PR 17-NOV-2000: 2000US-0249214.  
 PR 17-NOV-2000: 2000US-0249215.  
 PR 17-NOV-2000: 2000US-0249216.  
 PR 17-NOV-2000: 2000US-0249217.  
 PR 17-NOV-2000: 2000US-0249218.  
 PR 17-NOV-2000: 2000US-0249244.  
 PR 17-NOV-2000: 2000US-0249245.  
 PR 17-NOV-2000: 2000US-0249264.  
 PR 17-NOV-2000: 2000US-0249265.  
 PR 17-NOV-2000: 2000US-0249297.  
 PR 17-NOV-2000: 2000US-0249299.  
 PR 17-NOV-2000: 2000US-0249300.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 06-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 11-DEC-2000: 2000US-0251990.  
 PR 05-JAN-2001: 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI: 2001-483426/52.  
 XX  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 XX Disclosure: SEQ ID NO 39262; 3071pp + Sequence Listing; English.  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC to AAK87694 represent human immune/hematopoietic-derived cells. AAK64703  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 5271 BP; 1866 A; 897 C; 951 G; 1757 T; 0 other;  
 Alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 Alignment block:  
 US-08-973-363-6 x AAK84450 ..  
 Align seg 1/1 to: AAK84450 from: 1 to: 5271  
 27 SerArgAspLeuAlaLysArgIleu 34  
 ||||||||||||||||||||||||  
 4307 TCTAGGAGCTCTAGCAAGCGAGG 4330

---

OM of: US-08-973-363-6 to: EST:\* out\_format : pfs  
Date: Aug 3, 2002 5:46 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODL=framedp2n.model -DEV=rlh  
-O=/usr2/\_usr2/spool/US08973363/runat\_01082002\_080122\_19814/app-query.fasta\_1.638  
-DB=EST -QPR=fastap -SUFFIX=olip2n.rst -GAP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LDOPEL=0.000 -LDOPEXT=0.000  
-GAPOP=4.000 -GAPEXT=0.050 -XGAP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -XGAP=60.000 -XGAPEXT=60.000  
-DELDP=6.000 -DELDP=7.000 -START=1 -MATRIX=olip2n  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=plis -NORM=ext  
-HEAPSIZE=900 -MINLEN=0 -MAXLEN=200000000  
-USER=US08973363 -CGEN1\_1\_2938 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-08-973-363-6  
Query length: 41  
Database: EST:\*  
Database sequences: 13736207  
Database length: 184157050  
Search time (sec): 4311.51000

WARN: XGAPOP and XGAPEXT must be equal. Assuming XGAPOP=XGAPEXT=60.000  
WARN: XGAPEXT and XGAPEXT must be equal. Assuming XGAPEXT=XGAPEXT=60.000

Score list:

Sequence	Strid	Orig	ZScore	EScore	Len	Document
gb.esl1:BM155356	+	41.00	786.43	1.5e-34	619	BM155356 BH155356 RIKEN full-16
gb.esl1:BM161065	+	41.00	785.85	1.6e-34	660	BM161065 BM161065 RIKEN full-16
gb.esl1:BM161246	+	32.00	614.38	5.7e-25	430	BM161246 BM161246 RIKEN full-16
gb.esl1:BM190775	+	32.00	612.36	7.3e-25	547	BM190775 BM190775 RIKEN full-16
gb.esl1:BM25712	+	32.00	609.50	1.2e-24	866	BM25712 BM25712 RIKEN full-16
gb.esl1:BM25967	+	32.00	607.07	1.4e-24	1028	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	27.00	511.67	3.0e-19	821	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	26.00	494.94	2.5e-18	593	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	26.00	474.27	3.6e-18	645	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	22.00	441.31	2.5e-15	337	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	22.00	419.66	4.0e-14	438	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	18.00	336.22	1.8e-09	856	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	196.44	0.1075	129	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	191.13	0.2125	243	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	190.72	0.2238	255	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	189.64	0.2570	290	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	189.36	0.2665	300	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	188.46	0.2992	334	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	188.34	0.3040	339	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	187.92	0.3204	366	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	187.69	0.3301	355	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	187.27	0.3461	385	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.88	0.3661	403	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.66	0.3769	414	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.54	0.3828	420	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.44	0.3985	436	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.11	0.4044	442	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	186.03	0.4083	446	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	185.65	0.4320	470	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	185.25	0.4518	490	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	184.34	0.5076	546	BM25967 BM25967 RIKEN full-16
gb.esl1:BM25967	+	10.00	182.86	0.6295	667	BM25967 BM25967 RIKEN full-16

gb.gss:BM566359	+	9.00	161.72	9.23	796	BM566359 B08C9595F B08C Bras
gb.esl1:AM422113	+	8.00	150.01	41.44	316	AM422113 F157B02.Y1 Sugano K
gb.esl1:BM994052	+	8.00	148.60	49.68	374	BM994052 PM0-HR0913-120201-0
gb.esl1:BM999049	+	8.00	148.44	50.68	381	BM999049 PM0-HR0913-250401-0
gb.esl1:AM427958	+	8.00	148.40	50.97	383	AM427958 Fx50901.Y1 Soares T
gb.gss:AM265584	+	8.00	147.99	53.69	402	AM265584 RRC11-73H11 TK RCG
gb.esl1:BM447767	+	8.00	147.01	54.98	411	BM447767 G08H1903.X1 NICHAD X

seq\_name: gb.esl1:BM155356

seq\_documentation\_block:

LOCUS BM155356 619 bp mRNA linear EST 18-OCT-2001  
DEFINITION Musculus cDNA clone A130024116 3' similar to L10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
Accession  
BM155356  
Version  
BM155356.2 GI:16268254  
Keywords  
EST.  
Source  
Mus musculus.  
Organism  
house mouse.

REFERENCE

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanaoka,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koude,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyata,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Jun 28, 2000 this sequence version replaced gi:8811286.  
Contract: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center/GSC, Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-22 Saitama-shi, Tsukuba-shi, Ibaraki, Ibaraki, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc@riken.go.jp  
URL: http://genome-resgsc.riken.go.jp/  
Carninci,P., Shibata,K., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wegli,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384 format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Konno,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse cDNAs.  
Location/Qualifiers  
1. 619

FEATURES  
source





17 rargalaasptyrleuileylsleuleuseraraspaleuhalays 32  
|||||  
328 CCGTGACAGACTACCTCATCAATTAATTACTAGTAGAGATCTTCACAAA 283

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.

ACCESSION AU125712

VERSION AU125712.1 GI:10950428

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 866)

REFERENCE Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and

AUTHORS Isogai,T.

TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki

JOURNAL Y., Sugano,S., Isogai,T.)  
Unpublished (2000)

COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp

HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute: cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers

FEATURES  
source 1..866

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RM4002061"

/clone\_lib="NT2RM4"

/cell\_type="teratocarcinoma"

/cell\_line="NT2"

/note="Vector: PME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

alignment\_scores:

Quality: 32.00 Length: 32

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 lleleuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
450 ATTCTTCAGATGATCCGATATAAAACCAACCAACCAAGTTGCAGAC 499  
17 rargalaasptyrleuileylsleuleuseraraspaleuhalays 32  
|||||  
500 CCGTGACAGACTACCTCATCAATTAATTACTAGTAGAGATCTTCACAAA 545  
seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',

mRNA sequence.

ACCESSION BE895133

VERSION BE895133.1 GI:10358221

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1028)

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC/DCPD/DPF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.jnl.gov>  
Plate: LLM9753 Row: h Column: 16  
High quality sequence stop: 488.

FEATURES  
source 1..1028

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3921087"

/clone\_lib="NIH\_MGC-72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 387 a 205 c 238 g 198 t

ORIGIN

alignment\_scores:

Quality: 32.00 Length: 32

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 lleleuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
131 ATTCTTCAGATGATCCGATATAAAACCAACCAACCAAGTTGCAGAC 180  
17 rargalaasptyrleuileylsleuleuseraraspaleuhalays 32  
|||||  
181 CCGTGACAGACTACCTCATCAATTAATTACTAGTAGAGATCTTCACAAA 226  
seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
mRNA sequence.

ACCESSION BF239967

VERSION BF239967.1 GI:11153890

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 821)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE



JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9abbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM103 row: k column: 18  
High quality sequence stop: 562.

FEATURES  
source

1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="413129"  
/clone\_id="NH\_MGC\_34"  
/tissue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: bone marrow; Vector: pDMR-LIB (Clontech); Site\_1: SfiI (ggccgcttggcc); Site\_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGACGCGAGCGCGCAGCANG-CT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x BF239967 ..

Align seg 1/1 to: BF239967 from: 1 to: 821

6 Probasplyslyspgoclnalalysglnleuglntharqalaastrytie 22  
|||||  
3 CCCGTAATAAAACACACAGCAACAGTTCAGACCCGTCGAGACTACT 52  
|||||  
22 uilelysleuuserarqaspleuualys 32  
|||||  
53 CATCAATTACTAGTAGAGACTTGCACAAA 83

seq\_name: gb\_est1:AL659353

## seq\_documentation\_block:

LOCUS AL659353 593 bp mRNA linear EST 13-DEC-2001  
DEFINITION AL659353 XGC-neurula Silurana tropicalis cDNA clone TNeu045e20 5',  
mRNA sequence.

ACCESSION AL659353

VERSION AL659353.1 GI:17672995

KEYWORDS EST

SOURCE western clawed frog,  
Silurana tropicalis

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Silurana.

REFERENCE 1. (bases 1 to 593)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.

AUTHORS Sanger Xenopus tropicalis EST project 2001 (10\_2001)

TITLE Unpublished (2001)

JOURNAL Contact: Huckle E

COMMENT Sanger Centre

FEATURES  
source

Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropes@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TNeu045e20.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
Location/Qualifiers  
1. 593  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone\_image="TNeu045e20"  
/clone\_id="XGC\_neurula"  
/dev\_stage="neurula"  
/lab\_host="Escherichia coli DH10B"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dT primed from 5' end of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT 119 a 134 c 114 g 226 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x AL659353/rev ..

Align seg 1/1 to reverse of: AL659353 from: 1 to: 593

1 IileupProaspaspProaspplyslyspgoclnalalysglnleuglnth 17  
|||||  
346 ATTTTACGACGATGATCCAGATAGAAAGCCCAACGACGACTGACGAC 257  
|||||  
17 TATGAAAT 26  
|||||  
296 CAGAGCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 269

seq\_name: gb\_est1:AL644594

## seq\_documentation\_block:

LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001  
DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA  
sequence.

ACCESSION AL644594

VERSION AL644594.1 GI:16796719

KEYWORDS EST

SOURCE western clawed frog,  
Silurana tropicalis

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Silurana.

REFERENCE 1. (bases 1 to 645)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.

AUTHORS Sanger Xenopus tropicalis EST project 2001 (10\_2001)

TITLE Unpublished (2001)

JOURNAL Contact: Huckle E

COMMENT Sanger Centre

FEATURES  
source

Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropes@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: L1E1d12.pic  
Sequencing primer: P1C  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
Location/Qualifiers  
1. 645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone\_image="L1E1d12"

```
/clone_1lb="Xgc-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; CDNA
was oligo dT primed from 5ug of poly A+ RNA from egg
ECORI-NotI cut CDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

BASE COUNT      222 a 125 c 156 g 141 t 1 others
ORIGIN

alignment_scores:
  Quality:      26.00      Length:      26
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AL644594 ..

Align seg 1/1 to: AL644594 from: 1 to: 645

1 lileuproaspaspProaspLysProglInAlaLysGlnLeuGlnTh 17
|||||
456 ATTTTACGACATGATCCGATAGAGACCCCAAGCCAGCAGCTACAGAC 505
|||||
17 rArgAlaAspTyrIleuIleuLysLeuLeu 26
|||||
506 CAGAGCTGACTACCTCATTTAACTTCTC 533
|||||

seq_name: gb_est1:AM97058

seq_documentation_block:
LOCUS      AM97058          686 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens CDNA, mRNA sequence.
ACCESSION  AM97058
VERSION    AM97058.1 GI:8257292
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 686)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150
            400-152-c03&ct=2000-04-15&ft=1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 678.
            Location/Qualifiers
                1..686
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1lb="BN0047"
                /dev_stage="Adult"
```

```
/note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESPEs PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and CDNA amplification were performed under
low stringency conditions."

BASE COUNT      165 a 154 c 126 g 241 t
ORIGIN

alignment_scores:
  Quality:      25.00      Length:      25
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

9 lvsProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrIleuLysIle 25
|||||
104 AAACCAAGCAAGCAAAAGTTCAGACCCGTGCAGACGCTCATCAATTT 55
|||||
25 ulseuSerArgAspLeuAlaLysArg 33
|||||
54 ACTTAGTAGAGATCTTCGCAAAAAGA 30
|||||

seq_name: gb_est1:AM996787

seq_documentation_block:
LOCUS      AM996787          337 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION OV3-BN0047-230200-102-d03 BN0047 Homo sapiens CDNA, mRNA sequence.
ACCESSION  AM996787
VERSION    AM996787.1 GI:8257021
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 337)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230
            200-102-d03&ct=2000-02-23&ft=1)
            Seq primer: puc 18 forward
            High quality sequence start: 2
            High quality sequence stop: 337.
            Location/Qualifiers
                1..337
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1lb="BN0047"
                /dev_stage="Adult"
```



Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsic.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

FEATURES  
 source  
 1. 446  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="G930033j21"  
 /clone\_lib="RIKEN full-length enriched, mammary gland  
 RCB-0527 Jy9-MC(B) cDNA"  
 /tissue\_type="mammary gland"  
 /cell\_line="RCB-0527 Jy9-MC(B)"  
 BASE COUNT 139 a 99 c 108 g 100 t  
 ORIGIN

alignment\_scores:  
 Quality: 22.00 Length: 22  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x BB834922 ..

Align seg 1/1 to: BB834922 from: 1 to: 446

1 l1leuProAspPProAspLysPProGlnAlaLysGlnLeuGlnThr 17  
 |||||  
 309 ATTCTTCACATGATCCTGATMAAACACACAAACAGTTACAGAC 358  
 |||||  
 17 FARGAlAsPTyrlen 22  
 |||||  
 359 CCGTGCAGACTACTC 374

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:

LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigriviridis genome survey sequence 17 end of clone  
 103008 of library G from Tetradon nigriviridis, genomic survey  
 sequence.

ACCESSION AL286261 GI:8024707  
 VERSION AL286261.1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigriviridis.  
 ORGANISM Tetradon nigriviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 856)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,

Bonneau, L., Billault, A., Quetier, F., Saurin, M., Bernot, A. and  
 Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigriviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,  
 Bernot, A., Fizames, C., Winkler, P., Brothier, P., Quetier, F.,  
 Saurin, M. and Weissenbach, J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigriviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 REFERENCE  
 Genoscope.  
 Direct Submision  
 Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the tetradon nigriviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
 source  
 1. 856  
 /organism="Tetradon nigriviridis"  
 /db\_xref="taxon:99883"  
 /clone="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0B6103BF04LPI-end : T7"  
 BASE COUNT 84 a 308 c 252 g 206 t 6 others  
 ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 leuProAspPProAspLysPProGlnAlaLysGlnLeuGlnThr 18  
 |||||  
 674 CTGCCGATGATCCTGATMAAGCCCTCAGGCCAAGAGTTACAGACCAG 625  
 |||||  
 18 gAla 19  
 |||||  
 624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:

LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PM0-ET0208-031200-001-f12 ET0208 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M. R.,  
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,  
 Brinstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare  
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
 Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE  
 20202663  
 COMMENT  
 Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL:  
(http://www.ludwig.org.br/scripts/gethtml2.pl?L-PW06t2-PW0-ET0208-  
031200-001-f12xt3-2000-12-03&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 126.  
Location/Qualifiers

# FEATURES

## source

1. 129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung-tumor; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN

## alignment\_scores:

Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 lysginleuglnthrrgalaasptyleu 22  
|||||  
108 AAGCAGCTACAGACCGAGCGGATTCCTG 79

**THIS PAGE BLANK (USPTO)**



COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
source 1..6608 /organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN  
alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608  
1 lleleuproaspaspProaspPlysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAACCCAGCAAGCCTACAGAC 4129  
17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAAGCGCTTGCTGTGCA 4202  
seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chemo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cvs.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM  
Chicken.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source  
Location/Qualifiers  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-2"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/translation="MNGHSDSESVNRSSSESRDDSGSAGSGSGSSGSSGSSGSS  
SOGSSDSESGSESGSESDTSREKQYAKPKADGSEFWKSSPSIIAVQRSAY

LKQOQOQAAASDSSGSESSSEDSADSSSETKKKKKHNDQMOMSGSGSVSGTGS  
DSEADGDKSSCESESDYEPKRNKVKSKPPSRILPKSGKSTGQKKQOLDSSEEE  
DDDEDYDKRSRROATVNVSYKAEETKTDSDLLVEGCEDEVPOREDEFTETKEMD  
SRTGRGASTTITVAVPADDDPNAGDPKSKSEIEIOYLITWKGSHHNTWTEET  
LQOQVKGKNNKIDNTKKKQDETRKRLKNAKSPEDVEYINQOQLTDLHKQYIVERIT  
AHSNOKSAGYIPDYTCRWGDLPSSECSWEDGALIAKFFQARIDETFSRNQSTTPPKD  
CIVLKORPREVALAKROPYSYIGHESLELDYQNLNMLASHMCKGNSCLADENGSLG  
KTIOISFLNLYLFEHQLPPLRLRYPLSTLTMOWRELOTNAPOMNAVYLDITSRN  
MIRTHEMMHPTKRLKENLITTYETILKOKSFLGLNAPFVGEAHLKNDLSLY  
PILIDKSNHRLITGTPLONSLKELMSLHITPEKSSMEDFEHNGKREYAS  
LKELEPELILRRYKKDVEKSLPAKYEQILRMMSALOKOYIWTITRYNALKSKSGS  
STSGFLTMELKCCNHCYLILPPDNEFYIKOELQHLISSGKLILDLILRL  
EKGNYLIFSQWRLMIDIAEYLKROPPRLDGSIKELKROALDHFNAGSBDPC  
FLLSFRAGLGINLASADTVVIFSDMDNPNOLQOARAHRIQKQOYVIRLVYKGS  
VEEDILERRAKKKKMDHLYIORMDTTGKVTLTGSPSSSPFNKEESAILKPGAE  
LKEPGEPEPOEMDIDILIKRAETRENEPGLVYGDELLSQFVANFNSMDEIDIE  
LEPERNSRMEETIPESORRIIEBERQKLEETIMLPBRNCAKQISNGSEGRSR  
SRRYSGSDSDSITERRKRPKRPRPTIPRENTKGSDEIRFRIKSYKFGSLERLD  
AVARDAELVDKSETDLRLGELVHNGCALKADNSGGERAGRGKVGKPGPRJSGV  
QVNAKLIVISHEELAPLHKISIPDEERKRYIPCHTKAHHIDIMGKEDSDNLVGI  
YVYGGSWEMIKMDPLSLTOKILPDDPKPKOAKOLOTRADYLIKLKLDAREAO  
RLAGNSKRRKTRKNNKMAKSKIKETIKSPSPSPSKSPDEDEEDNKDEIYSVK  
HAKKITETKENEBKPEPDITGKKEAEKREKREKREKREKREKREKREKREKREK  
NKEKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK  
ESEELQKTFVSKCKEMRPVKALQOLBERGELSERDQLEHTROCLIKTIGHITFCL  
KEYTDEQIKQWRKMLMIFVSKFTEFDARKLYKHAIKKROESQOHDONMISNVN  
TVIRNPDIYERLEKTTNHDSSRDSYSDRHLKXOYDHDKHODGDAYKSPSRPY  
SAFNGKHDRMDYKODSRYSYDKKHKRLDHRDRHNSNEGMLKRSRSHDRSH  
SPDRHSDHRSSTSEYSHKSSRDYRHSMDONDHAASGSGSPRPIDQSRPYSRPLG  
HRSPEHSSDHRKSTPEHTWSKRT"  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872  
1 lleleuproaspaspProaspPlysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAACCCAGCAAGCCTACAGAC 4129  
17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAAGCGCTTGCTGTGCA 4202  
seq\_name: gb\_pat:A58684  
seq\_documentation\_block:  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN



BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
source 1..153  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58684 ..  
Align seg 1/1 to: A58684 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGAGCAAGCAGCTACAGACCCGTCGACGACTCATTA 101  
|||||  
24 sleuleuAsnLysAspLeuAlaArgLysGluAlaGlnArgLeuAlaGly 41  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGCAAGCAAGGCTTCTGTG 151  
41 la 41  
152 CA 153

seq\_name: gb\_pat:A58685

seq\_documentation\_block:  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58685 ..  
Align seg 1/1 to: A58685 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGAGCTACAGCTTACAGACCCGTCGACGACTCATTA 101  
|||||

24 sleuleuAsnLysAspLeuAlaArgLysGluAlaGlnArgLeuAlaGly 41  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGCAAGCAAGGCTTCTGTG 151  
41 la 41  
152 CA 153

seq\_name: gb\_pat:A58686

seq\_documentation\_block:  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58686 ..  
Align seg 1/1 to: A58686 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGAGCAAGCAGCTGACGACCCGTCGACGACTCATTA 101  
|||||  
24 sleuleuAsnLysAspLeuAlaArgLysGlu 34  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGAA 132

seq\_name: gb\_ro:MUSCHD1X

seq\_documentation\_block:  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Delmas, V., Stokes, D.G. and Perry, R.P.  
TITLE A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helix domain  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
MEDLINE 93211972  
REFERENCE 2 (bases 1 to 5349)  
AUTHORS Perry, R.P.

TITLE Direct Submission  
JOURNAL Submitted (08-Apr-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES  
source  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDSESVNNGSGESSQSDDCGSASGSGSGSSGSSGSS  
SOSGSDSDSGSDSGSDSESDTSREKNVQAKPKPVGCAEFKWSPPSLAVQRA  
RKQPOAQOQPASSNSGSEEDSSSSGSAKRKKHNDLGMQSSGSPQLG  
SDSESEERDRKSSCGTESDYEPKPKVSRKPNKSKNGKKILGOKRKQIDSE  
DDEDYDNDRSSRQATVNVSYKDEEMKTDSDLLVEYGEDVPQPEDEPPTIR  
DCRVGRKATGATTTTAAVEADGDPNAGPERKKEPDIOYLTKKGMSHIHNTMET  
TLKOONVRGAKKIDNKKKQDPTKMLKNASPEDEYTYNCOEILDDLHKQYI  
IAHSQKSAAGLPDYCKMGLPISECSWEDALISKFTQCTIDEYFSHNSKTY  
DCKVLKQRPREFVALKRPSTYIGHGELRDYQLNGLMWLASHWCKNGNSCLAD  
GKTQITQISFLNLFHEHQLYGFLLVPLSTLTSQREIQTASQMANAVYAGDINS  
NMIRTHEMMHPQTKRLKFNILTYEILTKDAFGLMMAFIVGEAHRKLNDSIL  
KTLIDFKSNHRLITGTPLONSLKEIMSLHFIEMKPSMEDFEHSGKREYGA  
SLHKELEPFLIRVKKDYKSLPAVEOILREMSALOKOYKATILTRYKALS  
GSTSGFLINMELKCCNHCYILKPPDNNEFNKODLHILRSSGKILDLKILRL  
REGRNVLIFSQVRMLDLIAEYLKRFQEPQRLDGSIGELRKQALDHPNAGSE  
GFLSTRAGGLGINLASADTVIVFDSDMNPQDLQAARHRIQOKOVNIYRLVY  
SVEEDILERRAKKMYLDLVIORMDTGKTVLHTGAPSSSTPFNKEELSAIKFG  
ELEKPEEGEOPQEMDIDELIKRAETHNEPGLTVGDELLSQFVANSNDE  
ELEPERNSKNMEIIPEDORRRLEEBEROKLEIETMLPRMNCAKOISFNS  
RSRRYSGSDSDSISERRKPKRGRPTIPRENTKFSDAEIRFTKSTKRGEP  
DAIARDAELVHDELIPLKSIIPSDPERKQYITPCHTKAAHFDIDWCKEDSN  
IYEVYGGSWEMIKMPDLSLTHKILPDDPKKPOAKLOTADYDILKLSRLAK  
OILGAGSGSKRRKTRAKSKAKMSIKVKEEIKSDSPLSEKSDDDDKINDS  
DRSKSVSDAIVHTTASGEPTVAESEBELDQKTFSTCKEMRPYKALAKOLD  
GLSEREQLHTRQCLIKTGDHTTECLAKESNPEOLKOWRKNLMTIVSKFTE  
KLYKHAIRKQESQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS  
QYHDKRQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS  
HRSRHRSLKGLKDRCHSDHRSRSHDHRSHDHRSHDHRSHDHRSHDHRSHD  
HRASSGSPSPIDORSYGRSPFEHSHRSTPEHTWSSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x MUSCHDIX ..  
Align seg 1/1 to: MUSCHDIX from: 1 to: 5349

1 TlaleuProaPaapProaPlySLySProGlnAlaLySglnleuGlnrh 17  
|||||  
4026 ATCTCTCCAGATGATCTGATATAAAAAACCAAGCAAGAGTATACGAC 4075  
17 TARGAlaAsPTyrlleuTleuTleu 26  
|||||  
4076 CCGTGCAGACTACTCTCAACTACTT 4103

seq\_name: gb\_pr:Af006513  
seq\_documentation\_block:  
LOCUS Af006513 5947 bp mRNA linear PRI 27-NOV-1997

DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 5947)  
Woodage,T., Basrai,M.A., Baxevanis,A.D., Hieter,P. and Collins,F.S.  
JOURNAL  
TITLE  
Characterization of the CHD family of proteins  
Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE  
97470991  
2 (bases 1 to 5947)  
REFERENCE  
AUTHORS  
Woodage,T.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
FEATURES  
source  
1..5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WI-5811"  
1..5947  
/gene="CHD1"  
164..5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db\_xref="GI:2645429"  
/translation="MNGHSDSESVNNGSGESSQSDDCGSASGSGSGSSGSSGSS  
SOSGSDSDSGSDSGSDSESDTSREKNVQAKPKPVGCAEFKWSPPSLAVQRA  
RKQPOAQOQPASSNSGSEEDSSSSGSAKRKKHNDLGMQSSGSPQLG  
SDSESEERDRKSSCGTESDYEPKPKVSRKPNKSKNGKKILGOKRKQIDSE  
DDEDYDNDRSSRQATVNVSYKDEEMKTDSDLLVEYGEDVPQPEDEPPTIR  
FMDCKIGRKATGATTTTAAVEADGDPNAGPERKKEPDIOYLTKKGMSHIHNTMET  
EETLKOONVRGAKKIDNKKKQDPTKMLKNASPEDEYTYNCOEILDDLHKQYI  
GKTQITQISFLNLFHEHQLYGFLLVPLSTLTSQREIQTASQMANAVYAGDINS  
NMIRTHEMMHPQTKRLKFNILTYEILTKDAFGLMMAFIVGEAHRKLNDSIL  
KTLIDFKSNHRLITGTPLONSLKEIMSLHFIEMKPSMEDFEHSGKREYGA  
SLHKELEPFLIRVKKDYKSLPAVEOILREMSALOKOYKATILTRYKALS  
FDCVCYLKQRPREFVALKRPSTYIGHGELRDYQLNGLMWLASHWCKNGNSCLAD  
GKTQITQISFLNLFHEHQLYGFLLVPLSTLTSQREIQTASQMANAVYAGDINS  
NMIRTHEMMHPQTKRLKFNILTYEILTKDAFGLMMAFIVGEAHRKLNDSIL  
KTLIDFKSNHRLITGTPLONSLKEIMSLHFIEMKPSMEDFEHSGKREYGA  
SLHKELEPFLIRVKKDYKSLPAVEOILREMSALOKOYKATILTRYKALS  
SGSTSGFLINMELKCCNHCYILKPPDNNEFNKODLHILRSSGKILDLKILRL  
REGRNVLIFSQVRMLDLIAEYLKRFQEPQRLDGSIGELRKQALDHPNAGSE  
GFLSTRAGGLGINLASADTVIVFDSDMNPQDLQAARHRIQOKOVNIYRLVY  
SVEEDILERRAKKMYLDLVIORMDTGKTVLHTGAPSSSTPFNKEELSAIKFG  
KGYVEDILERRAKKMYLDLVIORMDTGKTVLHTGAPSSSTPFNKEELSAIKFG  
AELEKPEEGEOPQEMDIDELIKRAETHNEPGLTVGDELLSQFVANSNDE  
DIELEPERNSKNMEIIPEDORRRLEEBEROKLEIETMLPRMNCAKOISFNS  
RSRRYSGSDSDSISERRKPKRGRPTIPRENTKFSDAEIRFTKSTKRGEP  
RDLATARPALVYKSETDRLRIGELVHNGCALKLSDSSGTERGGRGAKKPTFRI  
SGVOVNAKLVLSHEBELIPLKSIIPSDPERKQYITPCHTKAAHFDIDWCKEDSN  
IYEVYGGSWEMIKMPDLSLTHKILPDDPKKPOAKLOTADYDILKLSRLAK  
EALSGAGSKRRKTRAKSKAKMSIKVKEEIKSDSPLSEKSDDDDKINDS  
DRSKSVSDAIVHTTASGEPTVAESEBELDQKTFSTCKEMRPYKALAKOLD  
GLSEREQLHTRQCLIKTGDHTTECLAKESNPEOLKOWRKNLMTIVSKFTE  
KLYKHAIRKQESQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS  
QYHDKRQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNSQNS  
HRSRHRSLKGLKDRCHSDHRSRSHDHRSHDHRSHDHRSHDHRSHDHRSHD  
RASSGSPSPIDORSYGRSPFEHSHRSTPEHTWSSRKT"

BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x MUSCHDIX ..

US-08-973-363-7 x AF006513

Align seg 1/1 to: AF006513 from: 1 to: 5947

1 HleauProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4025 ATTCTTCAGATGATCCGATATAAAACACACAAAGCAAAACAGTTGCAGAC 4074  
17 fargAlaAspTyrLeuLeuLysLeuLeu 26  
|||||  
4075 CCGTGCAGACTACCTCATCAATTAATCTT 4102

seq\_name: gb\_pr:AC092372

seq\_documentation\_block:

LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001

DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.

ACCESSION AC092372

VERSION AC092372.3 GI:17402768

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 101220)

TITLE DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Unpublished

2 (bases 1 to 101220)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 101220)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

On Dec 7, 2001 this sequence replaced gi:15290448.

Date Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu

Quality: Phrap Quality &gt;=40 100% of Sequence;

Estimated Total Number of Errors is 0.

NOTE: This insert is not the entire sequence of the clone (entire

sequence is 146,7kb). It is clipped at the overlap with AC012624.

The number of bases overlapped is 90404.

Location/Qualifiers

1. 101220

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-58M12"

BASE COUNT 34122 a 18862 c 17827 g 30409 t

ORIGIN

alignment\_scores:

Quality: 26.00 Length: 26

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignement\_block:

US-08-973-363-7 x AC092372/rev

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 HleauProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17

|||||

26932 ATTCTTCAGATGATCCGATATAAAACACACAAAGCAAAACAGTTGCAGAC 26883

17 fargAlaAspTyrLeuLeuLysLeuLeu 26

26882 CCGTGCAGACTACCTCATCAATTAATCTT 26855

seq\_name: gb\_pr:AC012624

seq\_documentation\_block:

LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001

DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.

ACCESSION AC012624

VERSION AC012624.6 GI:14993679

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 134365)

TITLE DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Unpublished

2 (bases 1 to 134365)

DOE Joint Genome Institute.

Direct Submission

Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 134365)

DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission

Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

On Jul 21, 2001 this sequence replaced gi:14277267.

Location/Qualifiers

1. 134365

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTD-2082117"

BASE COUNT 40414 a 24497 c 25503 g 43951 t

ORIGIN

alignment\_scores:

Quality: 26.00 Length: 26

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignement\_block:

US-08-973-363-7 x AC012624

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 HleauProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17

|||||

118247 ATTCTTCAGATGATCCGATATAAAACACACAAAGCAAAACAGTTGCAGAC 118296

17 fargAlaAspTyrLeuLeuLysLeuLeu 26

|||||

118297 CCGTGCAGACTACCTCATCAATTAATCTT 118324

seq\_name: gb\_htg:AC021449

seq\_documentation\_block:

LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000

DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered

pieces.

ACCESSION AC021449

VERSION AC021449.3 GI:10047806

KEYWORDS HTG: HTGS\_PHASE1, HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 143079)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Homo sapiens, clone Rpl1-58M12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 143079)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,  
 Chappel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,  
 Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J.,  
 Ferreira, S., Grant, G., Hagos, B., Hearford, A., Horton, L.,  
 Garfella, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,  
 Landers, J., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
 MacDonald, P., Margulis, N., McEwan, P., McGuirk, A., McKernan, K.,  
 McPherson, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, D.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,  
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
 Zimmer, A. and Zody, M.  
 DIRECT SUBMISSION  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:1407963.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5154  
 Center clone name: 58\_M.12  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 134743 bases at least Q40  
 Consensus quality: 139227 bases at least Q30  
 Consensus quality: 140814 bases at least Q20  
 Insert size: 144000; agarose-fp  
 Insert size: 142179; sum-of-coverage  
 Quality coverage: 4.6 in Q20 bases; sum-of-coverage  
 Quality coverage: 4.7 in Q20 bases; sum-of-coverage  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 38820: contig of 38820 bp in length  
 \* 38821 38920: gap of 100 bp  
 \* 38921 40411: contig of 1491 bp in length  
 \* 40412 40511: gap of 100 bp  
 \* 40512 43279: contig of 2768 bp in length  
 \* 43280 43379: gap of 100 bp  
 \* 43380 46905: contig of 3526 bp in length  
 \* 46906 47005: gap of 100 bp  
 \* 47006 51830: contig of 4825 bp in length  
 \* 51831 51930: gap of 100 bp  
 \* 51931 62619: contig of 10689 bp in length  
 \* 62620 62719: gap of 100 bp  
 \* 62720 75408: contig of 12689 bp in length

FEATURES  
 source  
 1..143079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Rpl1-58M12"  
 /clone\_lib="RPCR-11 Human Male BAC"  
 1..38820  
 /note="assembly-fragment  
 vector\_side:left"  
 clone\_end:SP6  
 misc\_feature  
 1..38820  
 /note="assembly-fragment"  
 40512..43279  
 /note="assembly-fragment"  
 43380..46905  
 /note="assembly-fragment"  
 47006..51830  
 /note="assembly-fragment"  
 51931..62619  
 /note="assembly-fragment"  
 62720..75408  
 /note="assembly-fragment"  
 75509..92516  
 /note="assembly-fragment"  
 92617..106409  
 /note="assembly-fragment"  
 106510..143079  
 /note="assembly-fragment  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-7 x AC021449 ..  
 Align seq 1/1 to: AC021449 from: 1 to: 143079  
 1 11leupProAspAspProAspLysLysProGlnAlaLysGlnLeuGln 17  
 116147 ATTCTTCAGATGATCCGATTAACCAACGACAAACAGTTGCAGAC 116196  
 17 rArgAlaAspTyrIleuIleLysLeu 26  
 116197 CCGTGCAGACTACCTCATCAATTACTT 116224  
 seq\_name: gb\_hlg:AC008531  
 seq\_documentation\_block:  
 LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
 7 ordered pieces.  
 ACCESSION AC008531  
 VERSION AC008531.3 GI:12830078  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 145659)  
 AUTHORS DOE Joint Genome Institute.



```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57460: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
    source          1..193446
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="RP11-36012"
                    /clone_11b="RP11 human BAC library 11"
BASE COUNT      57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-08-973-363-7 x AC091946 ..
Align seg 1/1 to: AC091946 from: 1 to: 193446
1 11leuProAspProAspLysProGlnAlaLysGlnLeuGlnth 17
|||||
87658 ATTCTTCCAGATGATCCGATATAAAACCAACGCAAAACAGTTGCAGAC 87707
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
87708 CCGTGCAGACTACCTCATCAATTACTT 87735
seq_name: gb_pr:AC026778
```

```
seq_documentation_block:
LOCUS      AC026778      195433 bp      DNA      linear      PRI 01-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.
ACCESSION  AC026778
VERSION    AC026778.4  GI:14277282
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 195433)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE      Direct Submission
            JOURNAL     Unpublished
REFERENCE  2 (bases 1 to 195433)
            DOE Joint Genome Institute.
            TITLE      Direct Submission
            JOURNAL     Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 195433)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE      Direct Submission
            JOURNAL     Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Jun 1, 2001 this sequence version replaced gi:13677045.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.igf.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            WI-13675 G33101
            SHGC-58345 G38487
            SHGC-103595 G57841.
FEATURES
    source          1..195433
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTC-428111"
BASE COUNT      62762 a 37302 c 37040 g 58329 t
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-08-973-363-7 x AC026778/rev ..
Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433
1 11leuProAspProAspLysProGlnAlaLysGlnLeuGlnth 17
|||||
29732 ATTCTTCCAGATGATCCGATATAAAACCAACGCAAAACAGTTGCAGAC 29683
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
29682 CCGTGCAGACTACCTCATCAATTACTT 29655
seq_name: gb_pr:AC022121
seq_documentation_block:
LOCUS      AC022121      219258 bp      DNA      linear      PRI 30-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION  AC022121
VERSION    AC022121.6  GI:15375145
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
2 (bases 1 to 219258)  
UNPUBLISHED  
DOE Joint Genome Institute.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
DIRECT SUBMISSION  
Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
DIRECT SUBMISSION  
Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced gi:15148108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
STS Content:  
WT-5811 G04974  
WT-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841

FEATURES

Source  
1..219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x AC022121/rev

Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258

1 HleauProaapapProaaplyslsProGlnAlaLysGlnLeuGlnth 17  
|||||  
89450 ATCTTCAGATGATCCGATGAAACCAACCAACAGTTCAGAC 89401  
17 rArgalaasPTyrleuIleLysleuLeu 26  
|||||  
89400 CCGTGCAGACTACCTCATCAATTAATT 89373

**THIS PAGE BLANK (USPTO)**





1188 TCGTGCAGATTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 1237  
 34 luAlaGlnArGleuAlaGlyAla 41  
 ||||||||||||||||||  
 1238 AACGACAGAGACTTGCTGCTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

XX AA142751;

XX 12-MAR-1997 (first entry)

XX Chicken CHD-1A gene.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KM CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

XX Gallus sp.

XX Key Location/Qualifiers

FT CDS 228..5390

FT /tag= a

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 1; Fig 5; 76pp; English.

XX The chicken CHD-W gene (AA142754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AA142751) is suggested to initiate

CC female development in birds. The sequence of CHD-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AA142755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-fertile

CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 41.00 Length: 41

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lIleuArProAspProAspLysLysProGlnAlaLysGlnLeuGlnH 17  
 ||||||||||||||||||  
 4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGACGCTACAGAC 4129

17 rArGAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
 ||||||||||||||||||  
 4130 CCGTGACAGACTACTCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
 34 luAlaGlnArGleuAlaGlyAla 41  
 ||||||||||||||||||  
 4180 AACGACAAAGGCTTGCTGCTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq\_documentation\_block:

ID AA142757 standard; DNA; 153 BP.

XX AA142757;

XX 12-MAR-1997 (first entry)

XX Chick CHD-1A gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KM CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

PT P-PSDB; AAM08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),

CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)

CC and the great tit CHD-W gene (see also AA142759). Translated amino

CC acid sequences of this region are provided in AAM08146-49. The

CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes

CC determine sex in birds and can be used to identify the sex of an

CC embryo, foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 34.00 Length: 34

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AA142757 ..

Align seg 1/1 to: AA142757 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCGACGCAAGACGCTACAGACCCGTCGAGCTCACTCATTTAA 101
|||||
24 sLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 41
|||||
102 ATTACTGATTAAGACCTTGCAAGAAAGACACAAAGGCTGCTGTG 151
|||||
41 la 41
||
152 CA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758

seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
AC AA142758:
XX
XX 12-MAR-1997 (first entry)
XX
DE Chick CHD-W gene fragment.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
XX Key Location/Qualifiers
XX FH misc_difference 52..81
XX FT /*tag= a
XX FT /note= "bases 52-81 are a repeat of bases 22-51
XX FT and are ignored in the translated amino
XX FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX PA (ISIS-) ISIS INNOVATION LTD.
XX
XX PI Griffiths R, Tiwari B;
XX DR WPI: 1997-043127/04.
XX DR P-PSDB; AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX CC and the great tit CHD-W gene (AA142759). Translated amino acid
XX CC sequences of this region are provided in AAM08146-49. The CHD-1A
XX CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX CC sex in birds and can be used to identify the sex of an embryo,
XX CC foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
us-08-973-363-7 x AA142758 ..
Align seg 1/1 to: AA142758 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCGACGCTAAGCAGTTACAGACCCGTCGAGATTACTCATTTAA 101
|||||
24 sLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 41
|||||
102 ATTACTGATTAAGACCTTGCAAGAAAGACACAAAGGCTGCTGTG 151
|||||
41 la 41
||
152 CA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
AC AA142759:
XX
XX 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
XX Key Location/Qualifiers
XX FH misc_difference 52..81
XX FT /*tag= a
XX FT /note= "bases 52-81 are a repeat of bases 22-51
XX FT and are ignored in the translated amino
XX FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX PA (ISIS-) ISIS INNOVATION LTD.
XX
XX PI Griffiths R, Tiwari B;
XX DR WPI: 1997-043127/04.
XX DR P-PSDB; AAM08149.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX PT birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX CC and the great tit CHD-W gene (AA142759). Translated amino acid
XX CC sequences of this region are provided in AAM08146-49. The CHD-1A
XX CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX CC sex in birds and can be used to identify the sex of an embryo,
XX CC foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLeu 24  
|||||  
52 AAGAAACCAAGGCAAGCAGTTGCAGACCCCTGCAGATTACTCATTTAA 101  
24 sLeuLeuAsnLysAspLeuAlaArgLysGlu 34  
|||||  
102 ATTACTGATATAAGACCTTGCAAGAAAGAA 132

seq\_name: /SIDSL/gcgdata/hold-geneseq/gene-seq-n-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; cDNA; 1311 BP.

AC AAV59280;

DT 14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

XX ds; human: telomere repeat binding factor: A-TRF: dimerisation domain;

KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

OS Homo sapiens.

OS Synthetic.

FH Key Location/Qualifiers

FT CDS 1..1311

FT /\*tag= a

FT /product= "A-TRF"

XX MO9836066-A1.

PD 20-AUG-1998.

PF 13-FEB-1998; 98WO-US02765.

PR 04-FEB-1998; 98US-0018628.

PR 13-FEB-1997; 97US-0800264.

XX (UTRQ ) UNIV ROCKEFELLER.

PI Bianchi A, De Lange T, Van Steensel B;

XX WPI: 1998-480769/41.

DR P-PSDB: AAM59280.

XX Nucleic acid encoding altered telomere repeat binding protein and  
PT related vectors - transformants, hetero-dimers and antibodies, used  
PT to inhibit shortening of telomerases caused by ageing or disease,  
PT also used to extend life of cells in culture

XX Claim 14; Page 110-111; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
CC inhibit shortening of telomeres associated with ageing (for cosmetic  
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
CC tumours and viral (including human immune deficiency virus) infection.  
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
CC for expression of recombinant proteins or where intended for subsequent  
CC transplant or for testing, eliminating the need for transformation.

XX S0 Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

alignment\_scores:

Quality: 25.00 Length: 25

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAV59280 ..

Align seg 1/1 to: AAV59280 from: 1 to: 1311

1 lIleLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThr 17  
|||||  
1237 ATTCTTCGAGATGATCCTGTATATAAACCAAGCAAGAAACAGTTACAGAC 1286  
17 rArgAlaAspTyrLeuIleLysLeu 25  
|||||  
1287 CCGTGCAGACTACTCATCAACTA 1311

seq\_name: /SIDSL/gcgdata/hold-geneseq/gene-seq-n-emb1/NA1997.DAT:AAT42756

seq\_documentation\_block:

ID AAT42756 standard; DNA; 153 BP.

AC AAT42756;

DT 12-MAR-1997 (first entry)

DE Mouse CHD-1 gene (bases 3855-977).

XX Bird; sex determination; chromodomain-helicase-DNA binding 1;

KW CHD-1; CHD-W; W chromosome; ss.

XX Mus sp.

FH Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

XX MO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

XX WPI: 1997-043127/04.

DR P-PSDB: AAW08146.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of

CC progeny.  
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
SQ

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

8 LysFrogInAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAAAACCCAAAGCAACAGTTCACGACCGGTGACAGACTCATCA 101  
24 sLeuLeu 26  
|||||  
102 ACTACTT 108

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:  
ID ABL06443 standard; cDNA: 6240 BP.  
XX  
AC ABL06443;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB; ABB62340.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 13811; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;  
SQ

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

9 LysFrogInAlaLysGlnLeuGlnThrArgAla 19  
|||||  
4205 AAGCCCGAGCCAGACAGCTGACGACCGGTGCC 4237

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:  
ID ABL06442 standard; cDNA: 9933 BP.  
XX  
AC ABL06442;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB; ABB62339.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 13808; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 9933 BP; 2711 A; 2460 C; 2538 G; 2216 T; 0 other;  
SQ

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:  
US-08-973-363-7 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

9 LysProGlnAlaLysGlnLeuGlnThrGala 19  
|||||  
6745 AACGCCAGCCAGCAGCTGTCAGACGCGTCCC 6777

seq\_name: /SIDSI/gcsgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882

seq\_documentation\_block:

ID AAK88882 standard; cDNA; 421 BP.

AC AAK88882;

DT 05-NOV-2001 (first entry)

DE Human digestive system antigen coding sequence SEQ ID NO: 1198.

KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
digestive system disorder; Meckel's diverticulum; ss.

OS Homo sapiens.

PN WO200155314-A2.

PD 02-AUG-2001.

PE 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.

```

XX PR 17-NOV-2000; 2000US-0249215.
XX PR 17-NOV-2000; 2000US-0249216.
XX PR 17-NOV-2000; 2000US-0249217.
XX PR 17-NOV-2000; 2000US-0249218.
XX PR 17-NOV-2000; 2000US-0249219.
XX PR 17-NOV-2000; 2000US-0249224.
XX PR 17-NOV-2000; 2000US-0249264.
XX PR 17-NOV-2000; 2000US-0249265.
XX PR 17-NOV-2000; 2000US-0249297.
XX PR 17-NOV-2000; 2000US-0249298.
XX PR 17-NOV-2000; 2000US-0249300.
XX PR 01-DEC-2000; 2000US-0250160.
XX PR 01-DEC-2000; 2000US-0250391.
XX PR 03-DEC-2000; 2000US-0251030.
XX PR 03-DEC-2000; 2000US-0251058.
XX PR 03-DEC-2000; 2000US-0256719.
XX PR 06-DEC-2000; 2000US-0251856.
XX PR 08-DEC-2000; 2000US-0251868.
XX PR 08-DEC-2000; 2000US-0251869.
XX PR 08-DEC-2000; 2000US-0251980.
XX PR 11-DEC-2000; 2000US-0254097.
XX PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM,
DR WPI: 2001-502630/55.
XX P-FSDB: AAM93109.
XX PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognostizing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX PS Claim 1: SEQ ID NO 1198; 986bp; English.
XX CC The present invention provides the protein and coding sequences of
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer. Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis o
utera rectal colitis. The present sequence is a cDNA encoding a diges
system antigen of the invention.
XX SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-7 x AAK88882 ..
Align seg 1/1 to: AAK88882 from: 1 to: 421
13 LysGlnIleuGlphrArAlaSPyrIleu 22
|||||
146 AACGACGTACAGACC CGACGCATTAATTG 175
seq_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn_embd1/NA2001A.DAT:AAI57603
seq_documentation_block:
ID AAI57603 standard; cDNA: 421 BP.
XX AC AAI57603;
XX DT 19-OCT-2001 (first entry)
XX DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.
```

XX	Human: colorectal cancer; colorectal cancer antigen; gene therapy; ss
XX	Homo sapiens.
XX	WO200155350-A1.
XX	02-AUG-2001.
XX	17-JAN-2001; 2001WO-US01350.
XX	31-JAN-2000; 2000US-0179065.
XX	04-FEB-2000; 2000US-0180628.
XX	24-FEB-2000; 2000US-0184664.
XX	02-MAR-2000; 2000US-0186350.
XX	16-MAR-2000; 2000US-0189874.
XX	17-MAR-2000; 2000US-0190076.
XX	18-APR-2000; 2000US-0198123.
XX	19-MAY-2000; 2000US-0205515.
XX	07-JUN-2000; 2000US-0209467.
XX	28-JUN-2000; 2000US-0214886.
XX	30-JUN-2000; 2000US-0215135.
XX	07-JUL-2000; 2000US-0216647.
XX	07-JUL-2000; 2000US-0216880.
XX	11-JUL-2000; 2000US-0217487.
XX	11-JUL-2000; 2000US-0218496.
XX	26-JUL-2000; 2000US-0220963.
XX	26-JUL-2000; 2000US-0220964.
XX	14-AUG-2000; 2000US-0224518.
XX	14-AUG-2000; 2000US-0224519.
XX	14-AUG-2000; 2000US-0225213.
XX	14-AUG-2000; 2000US-0225214.
XX	14-AUG-2000; 2000US-0225265.
XX	14-AUG-2000; 2000US-0225267.
XX	14-AUG-2000; 2000US-0225268.
XX	14-AUG-2000; 2000US-0225270.
XX	14-AUG-2000; 2000US-0225274.
XX	14-AUG-2000; 2000US-0225757.
XX	14-AUG-2000; 2000US-0225758.
XX	14-AUG-2000; 2000US-0225759.
XX	18-AUG-2000; 2000US-0226279.
XX	22-AUG-2000; 2000US-0226681.
XX	22-AUG-2000; 2000US-0226868.
XX	22-AUG-2000; 2000US-0227182.
XX	23-AUG-2000; 2000US-0227009.
XX	30-AUG-2000; 2000US-0228924.
XX	01-SEP-2000; 2000US-0229287.
XX	01-SEP-2000; 2000US-0229343.
XX	01-SEP-2000; 2000US-0229344.
XX	01-SEP-2000; 2000US-0229345.
XX	05-SEP-2000; 2000US-0229509.
XX	05-SEP-2000; 2000US-0229513.
XX	06-SEP-2000; 2000US-0230437.
XX	06-SEP-2000; 2000US-0230438.
XX	08-SEP-2000; 2000US-0231242.
XX	08-SEP-2000; 2000US-0231243.
XX	08-SEP-2000; 2000US-0231244.
XX	08-SEP-2000; 2000US-0231413.
XX	08-SEP-2000; 2000US-0231414.
XX	08-SEP-2000; 2000US-0232036.
XX	12-SEP-2000; 2000US-0232081.
XX	12-SEP-2000; 2000US-0232081.
XX	14-SEP-2000; 2000US-0232397.
XX	14-SEP-2000; 2000US-0232398.
XX	14-SEP-2000; 2000US-0232399.
XX	14-SEP-2000; 2000US-0232400.
XX	14-SEP-2000; 2000US-0232401.
XX	14-SEP-2000; 2000US-0233063.
XX	14-SEP-2000; 2000US-0233064.
XX	14-SEP-2000; 2000US-0233065.
XX	21-SEP-2000; 2000US-0234223.
XX	21-SEP-2000; 2000US-0234224.

PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-457727/49.  
DR P-PSDB: AAM38625.  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon and rectum including colorectal cancers  
PT and also for testing and detection e.g. diagnosis -  
XX  
XX  
PS Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in  
CC AAI57547, AAI57619 and AAM38569-AAM38641. These can be used in the  
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
CC The present sequence is a colorectal cancer antigen coding sequence of  
CC the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other:  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAI57603 ..  
  
Align seg 1/1 to: AAI57603 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
146 AAGCAGCTACAGACCGAGCGATCTACTTG 175  
  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV33521  
seq\_documentation\_block:  
ID AAV33521 standard; cDNA; 252 BP.  
XX  
AC AAV33521;  
XX  
DT 29-DEC-1998 (first entry)  
XX  
DE Clone 23789 cDNA fragment encoding a human protein kinase homolog.  
XX  
KW Protein kinase; cell signalling; inflammation; carcinoma; diabetes;  
KW human x-linked agammaglobulinemia; nonpheryocytic hemolytic anaemia;  
KW arteriosclerosis; glioma; restenosis; cholera-based septic shock;  
KW CHKRRNK chicken tyr kinase; ss.  
XX  
OS Homo sapiens.  
XX  
PN US5817479-A.  
XX  
PD 06-OCT-1998.  
XX  
PF 07-AUG-1996; 96US-0700575.  
XX  
PR 07-AUG-1996; 96US-0700575.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Au-Young J, Bandman O, Hawkins PR, Wilde CG;  
XX WPI: 1998-556387/47.  
DR



```
XX Human kinase poly(nucleotide(s) and recombinant products - useful
PT for identification of modulators of the enzyme, and treatment of
PT diseases associated with abnormal kinase expression
XX
PS Claim 1: Columns 29-30: 30pp; English.
XX
CC The invention provides polynucleotides which encode novel protein
CC kinase homologs expressed in various human cells and tissues. The
CC present sequence represents the clone 23789 cDNA fragment derived
CC from a human inflamed adenoid cDNA library. The cDNA encodes a
CC protein kinase which shows homology to the CKFRNK chicken tyr kinase.
CC Vectors and host cells can be used for recombinant production of the
CC protein kinase homolog. The recombinant proteins may be used to raise
CC antibodies for use as anti-kinase therapeutics. Oligonucleotides based
CC on the polynucleotide sequences, i.e. probes and antisense constructs,
CC the peptides and antibodies are claimed to be useful as tools for
CC studying signaling cascades in cells and proteins, and for identifying
CC inhibitors (drugs) to treat diseases and inflammatory conditions
CC associated with abnormal kinase expression. Diseases that are claimed
CC to be treatable include human X-linked agammaglobulinemia,
CC nonspherocytic haemolytic anaemia, arteriosclerosis, carcinomas,
CC diabetes, gliomas, restenosis, cholera-based septic shock, etc.
XX
SQ Sequence 252 BP; 63 A; 70 C; 72 G; 47 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AAV33521 ..
Align seg 1/1 to: AAV33521 from: 1 to: 252

24 LysleuLeuAsnLysAspLeuAla 31
|||||
100 AACTGCTCAACAAAGACCTGCA 123

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA61903

seq_documentation_block:
ID ABA61903 standard; DNA: 544 BP.
XX
AC ABA61903:
XX
DE 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #10208.
XX
KW Human; foetal liver; gene expression: single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
PS Claim 1: SEQ ID NO 10208; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x ABA61903 ..
Align seg 1/1 to: ABA61903 from: 1 to: 544

21 TyrleuileLysleuAsnLys 28
|||||
243 TACCTGATTAAGTATTAAATAAG 266

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK10214

seq_documentation_block:
ID AAK10214 standard; DNA: 544 BP.
XX
AC AAK10214:
XX
DE 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 10205.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
```


PT brains -  
XX  
PS Example 4; SEQ ID NO: 10205; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAK10214 ..  
  
Align seg 1/1 to: AAK10214 from: 1 to: 544  
  
21 TyrlleuilelystleuLeuAsnLys 28  
|||||  
243 TACCTGATAAGTTATTAATAAG 266  
  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK36112  
  
seq\_documentation\_block:  
ID AAK36112 standard; DNA; 544 BP.  
XX  
AC AAK36112;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR MPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
XX  
PS Example 4; SEQ ID NO: 10669; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAK36112 ..  
  
Align seg 1/1 to: AAK36112 from: 1 to: 544  
  
21 TyrlleuilelystleuLeuAsnLys 28  
|||||  
243 TACCTGATAAGTTATTAATAAG 266

Mon Aug 5 11:51:58 2002

us-08-973-363-7.clip2n.rng

Page 11



**THIS PAGE BLANK (USPTO)**



DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 645)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
AUTHORS Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
TITLE Unpublished (2001)  
JOURNAL Contact: Huckle E  
COMMENT Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: L1E1d12.pic  
Sequencing primer: PIC  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
FEATURES  
source  
1..645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="L1E1d12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"  
BASE COUNT 222 a 125 c 156 g 141 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 30.00 Length: 30  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x AL644594 ..  
Align seg 1/1 to: AL644594 from: 1 to: 645  
1 l1eleuproaspappproaspplyslpsproglalalysglnleuglnth 17  
|||||  
456 ATTTCACGATGATCCAGATTAAGAACCCCAAGCAGCTACAGAC 505  
17 rargalasptryleuilelysleuauanlyaspneu 30  
|||||  
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGATCTG 545  
seq\_name: gb\_est1:AL601246  
seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFZP313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKFZP313J1040 5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Weil,B. and Wismann,S.

TITLE EST (Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B. and Wismann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloecker H  
MIPS  
Am Klopfersplitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wismann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wismann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No sl sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
FEATURES  
source  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; site\_1: SfiI; site\_2: SfiIB; cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x AL601246 ..  
Align seg 1/1 to: AL601246 from: 1 to: 430  
1 l1eleuproaspappproaspplyslpsproglalalysglnleuglnth 17  
|||||  
218 ATTCTTCAGATGATCCGATTAAGAACCCCAAGCAGCTACAGAC 267  
17 rargalasptryleuilelysleuau 26  
|||||  
268 CCGTCGAGACTACCTCATCAATTACTT 295  
seq\_name: gb\_est1:A1890775  
seq\_documentation\_block:  
LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wmg5f11.x1 NCL\_CGAP\_U22 Homo sapiens cDNA clone IMAGE:2443725 3'  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; , mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cga@b3.rtiemail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.



1 l1leuProaspPProaspPlyLysProGlnAlaLysGlnleuGlnh 17  
|||||  
489 ATCTCTCCAGATGATCTCTGATTAATAAACCAAGCAAAACAGTTACAGAC 538  
17 rArgAlaSPtyrLeuIleLysLeu 26  
|||||  
539 CCGTGCAGACTACTCTCATCAACTACTT 566  
seq\_name: gp\_est1:BB461065  
seq\_documentation\_block:  
LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to l10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
ACCESSION BB461065  
VERSION BB461065.2 GI:16426612  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 660)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Jul 21, 2000 this sequence version replaced gi:9356558.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp.  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
FEATURES  
Source 1. 660  
Location/Qualifiers

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_1lb="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Sall1; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATTCTCGAGTTTAAATTAAATCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLX I."

BASE COUNT 211 a 130 c 168 g 151 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x BB461065 ..  
Align seg 1/1 to: BB461065 from: 1 to: 660  
1 l1leuProaspPProaspPlyLysProGlnAlaLysGlnleuGlnh 17  
|||||  
481 ATCTCTCCAGATGATCTCTGATTAATAAACCAAGCAAAACAGTTACAGAC 530  
17 rArgAlaSPtyrLeuIleLysLeu 26  
|||||  
531 CCGTGCAGACTACTCTCATCAACTACTT 558  
seq\_name: gp\_est1:AU125712  
seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,  
Y., Sugano,S., Isogai,T.)  
Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1533-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: [genomics@hri.co.jp](mailto:genomics@hri.co.jp)



HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES  
source  
Location/Qualifiers

1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_id="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 HleuProaspPaspProaspPlyspProglAlaLysGlnLeuGlnth 17  
|||||  
450 ATTCTCCAGATGATCCGATTAACCAACAGCAAGCAAGAGTGCAGAC 499  
17 rArgAlaaspPlyrleuileLysleu 26  
|||||  
500 CCGTCGACGACTACCTCATCAATTAATTCTT 527

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 6014360801 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.

ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1028)  
NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

plate: L14M9753 row: h column: 16  
High quality sequence stop: 488.  
Location/Qualifiers

FEATURES  
source  
1. 1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_id="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="D10B (phage-resistant)"  
/note="Organ: skin; Vector: PCMV-SPORT6; site\_1: Nct1;

site\_2: Sal1; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 HleuProaspPaspProaspPlyspProglAlaLysGlnLeuGlnth 17  
|||||  
131 ATTCTCCAGATGATCCGATTAACCAACAGCAAGCAAGAGTGCAGAC 180  
17 rArgAlaaspPlyrleuileLysleu 26  
|||||  
181 CCGTCGACGACTACCTCATCAATTAATTCTT 208

seq\_name: gb\_est1:AM996787

seq\_documentation\_block:

LOCUS AM996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM996787  
VERSION AM996787.1 GI:8257021  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 337)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Gadgil,M.A., da Silva,M.J.R., Zago,M.A., Bordin,S., Costa,F.F.,  
Gadgil,M.A., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=825-QV3-BN0047-230>)

Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 337.  
Location/Qualifiers

FEATURES  
source  
1. 337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="BN0047"  
/dev\_stage="Adult"  
/note="Organ: breast-normal; Vector: puc18; site\_1: Sma1;  
site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t  
ORIGIN

alignment\_scores:           Quality: 23.00           Length: 23  
                          Ratio: 1.000           Gaps: 0  
Percent Similarity: 100.000   Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AW96787/rev ..

Align seg 1/1 to reverse of: AW96787 from: 1 to: 337

1 ltleuproaspaspProaspLysPProGlnAlaLysGlnLeuGlnth 17  
|||||  
105 ATCTCTCCAGATGATGCCGATAAAAACCAAGCAAGCAGTTGCAGAC 56

17 rargAlaspTyrleu 23  
|||||  
55 CCGTGCAGACTATCTCATC 37

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730  
KEYWORDS BB830730.1 GI:17008973  
EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 438)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL Unpublished (2001)  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Waeli,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

FEATURES  
Location/Qualifiers  
Source 1. .438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930013K04"  
/clone\_11b="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT 138 a 96 c 108 g 96 t  
ORIGIN

alignment\_scores:           Quality: 22.00           Length: 22  
                          Ratio: 1.000           Gaps: 0  
Percent Similarity: 100.000   Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

1 ltleuproaspaspProaspLysPProGlnAlaLysGlnLeuGlnth 17  
|||||  
300 ATCTCTCCAGATGATGCCGATAAAAACCAAGCAAGCAGTTACAGAC 349

17 rargAlaspTyrleu 22  
|||||  
350 CCGTGCAGACTATCTCATC 365

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922.1 GI:17013165  
KEYWORDS BB834922.1  
EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 446)

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL Unpublished (2001)  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuda, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

## FEATURES

Location/Qualifiers

1..446  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="G93003321"  
 /clone\_lib="RIKEN full-length enriched, mammary gland  
 RCB-0527 Jy9-MC(B) cDNA"  
 /tissue\_type="mammary gland"  
 /cell\_line="RCB-0527 Jy9-MC(B)"  
 BASE COUNT 139 a 99 c 108 g 100 t  
 ORIGIN

## alignment\_scores:

Quality: 22.00 Length: 22  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-7 x BB834922

Align seg 1/1 to: BB834922 from: 1 to: 446

1 11leuPProAspAspProAspLysLysProGlnAlaLysGlnLeuGln 17  
 |||||  
 309 ATTCTTCCAGATGATCTCATTAATAAACACACAGCCAAACGATTACGAC 358  
 17 rArgAlaAspTyrLeu 22  
 |||||  
 359 CCGTCACACACTACTCTC 374

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
 DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 mRNA sequence.

ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Plate: LCM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers

## FEATURES

source

1..821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4133129"  
 /clone\_lib="NIH\_MGC\_54"  
 /tissue\_type="from chronic myelogenous leukemia"  
 /lab\_host="DH10B (11 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccatagcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCGGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."  
 BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN

## alignment\_scores:

Quality: 21.00 Length: 21  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-7 x BF239967

Align seg 1/1 to: BF239967 from: 1 to: 821

6 ProAspLysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22  
 |||||  
 3 CCGATTAATAAACACACAAACAGTTGACAGCCGTGCAGACTACTCT 52  
 22 uileLysLeuLeu 26  
 |||||  
 53 CATCAATTAATCTT 65

seq\_name: gb\_est1:AW97058

seq\_documentation\_block:

LOCUS AW97058 686 bp mRNA linear EST 05-JUN-2000  
 DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW97058  
 VERSION AW97058.1 GI:8257292

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Dias Neto, E., Garcia, Correa, R., Verjovski-Almeida, S., Bioness, M.R.,

Nagai, M.A., de Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and

Simpson, A.V.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 2020263

COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-qv3-BN0047-150>)  
 400-152-c03et3-2000-04-15et4-1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers

FEATURES  
 source  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
 165 a 154 c 126 g 241 t

ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-7 x AW997058/rev ..

Align seg 1/1 to reverse of: AW997058 from: 1 to: 686

9 LysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleIysLe 25  
 ||||||||||||||||||||||||||||||||||||||||||||  
 104 AAACCCACAAACAAACAGTTGCAGACCCGTGCAGACTACCTCAATTT 55  
 25 ulau 26  
 ||||  
 54 ACTT 51

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:  
 LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
 103K08 of library G from Tetradon nigroviridis, genomic survey  
 sequence.  
 AL286261.1 GI:8024707  
 AL286261.1 GSS: genome survey sequence.  
 KEYWORDS Tetradon nigroviridis.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetradon.  
 1 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fitzames,C., Wincker,P., Brotier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigroviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 Genoscope.  
 Direct Submission  
 Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.  
 Location/Qualifiers

FEATURES  
 source  
 1..856  
 /organism="Tetradon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BG103BF04LP1-end : T7"  
 BASE COUNT  
 84 a 308 c 252 g 206 t 6 others

ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-7 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 LeuProAspAspProAspLysLysProGlnAlaIysGlnLeuGlnThrArg 18  
 ||||||||||||||||||||||||||||||||||||||||||||  
 674 CTGCCGATGATCTCTATTAAGAAGCTCAGGCCAACAAGTTACAGACAG 625  
 18 ga1a 19  
 ||||  
 624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:  
 LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PM0-ET0208-031200-001-f12 ET0208 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.G.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=PM0&l2=PM0-ET0208-031200-001-fl12&l3=2000-12-03&l4=1)

Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 128.

Location/Qualifiers

FEATURES

source  
1..129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung tumor; Vector: puc18; Site: 1: Sma1;  
Site: 2: Sma1; A mini-library was made by cloning products  
derived from ORSPRES PCR (U S Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
108 AAGCAGCTACAGACCGAGGATTAAGTCTG 79

**THIS PAGE BLANK (USPTO)**



COMMENT other publication AU 5906996 961224.  
FEATURES  
source 1..6608 /organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AS8691 ..

Align seg 1/1 to: AS8691 from: 1 to: 6608

1 ltleuProaspPProaspPlysProGlnAlaLysGlnLeuGlnth 17  
ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 4129  
4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 4129

17 rargAlaaspyrLeuileLysleuLeuasnLysaspleuAlaargLysg 34  
|||||  
4130 CCGTGCAGACTACCTCATTAATTAAGTAAGACCTTGCAGAAAG 4179

34 luAlaGlnarGleuAlaGlyAla 41  
|||||  
4180 AAGCACAAAGGCTTGCTGCTGCA 4202

seq\_name: gb\_ov:AF004397

seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK

FEATURES  
source  
location/Qualifiers  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501846"  
SOGSSDSSESGESGSESSESDTSREKQVQAPKPADGSEFWKSSPSIIAYORSAY

LKQOQOAKAASDSGSEDSSEDSADSSSETRKKKHDEDMQMSGSGSVGTS  
DSEAEEDGKSSCESESESDYKRRKPPSRKPSKSGKSTGQKRLDSEEE  
DDDEDYDGRSROATVNAVYKKEAEETKTDSDLEVCAGDVPQTEDEFEETKED  
SRIGRKATGATSTTIVAEADGPNAGPEKSEIGETIOYLTKKMSHINTMETEET  
LKOQNVGAMKLDNYKKKDOETKRWLNKASPEVEYVNCQELTDDHKQYQYERTI  
AHSNOKSAAGPYDYCKWQGLPYSECSWEGALIAKFFQARIDYESRNOCKTTPED  
CKYLKQRPREVALAKKOPSYIGHSELELDYQQLGMLWLAHSMCKNSCLIADEMGIG  
KTIOITISFLYLFHEHQLGFLRVLSTLTMORIEIQTAPQMAVVLIGDITSN  
MIRTHEMHQOTKRLKFNILLTTEYLILKDKSPFLGGLNMAFIDEAHLKNDLSLY  
RTLIDFSNHRILITGTPILONSILKELSLHFTIMEPEKSSMEDEEERHKGREYAS  
LHKLEPFLRYKAKDYKSLPAKVEDITLKNEMALOKOYKATITPNYKALSKSGSG  
STSGFLNIMELKCNHCTILPDDNEFTYKQELQHLIRSGKILILDKLIRL  
ERGNRLVIFSQWRLMDILAEYLKYPQRLDGLSTIGELRKALDHFNAESSEDC  
FLSTRAGGLGINIASADYVIFEDSDNPNQDLQARARHIGOKQVNIYRLVTKS  
VEEDILERRAKKVLVDLVIORMDTGKTVLHTGSPSSTPFKELSAILKGAEL  
LFKEPEGEPOPEMDIDELIKRAETRENEGGLTVGDELISOKVAFNMDDIE  
LEPERNRNNEEIIPEQORRIIEEEROKLEETVMPKRNCAKQISFNSEBRER  
SRRTSGSDSDITERKRPKRGRPTIPREVIKGFSDAETIRFISTKTFGPLERD  
AVARDAELVDKSETDLRLGELVHNGCIRKALKNSSQERAGRLGVKQPTERTISG  
QVNAKLIVISHHEELAPLHKSIPSDPEERKRVIPCHTKAAHFDIDMCKEDSNLVGI  
YEVYGSWEMIKMPDLSLOKILPDDPKPKQAKQOTRADYILKLNKDLARKQ  
RLAGAGNSKRRTKRNKKNKASKIKKEIKSDSSPOSEKSEDEDEEDNKDELVSX  
HLKKITTEKENEPEPDICIRKEAREKETKRNKRLKREKKEKDELKED  
NKEKRNKVESTOKEKEVEBEKVENKSEKESKIPILDTPVHTTATSEPISE  
ESELHOKTFESVCKERNRPVKAALQODREPGLSEBQLEHTRQCLIKIGDITTEL  
KEYNPQIKQWRKNLWIFVSKFTEPDARKLHLKYKAIRKROSOQHNQNSNVN  
THVIRNPDVERLKETTNHDSRDSYSSDRILSOYHDHNRGDGAYKRSDSKRPY  
SAPNSGDNDRMDHYKODSRYSDSKRKLDDHRSRHSRHSLEGNLKDNGHSDHRS  
SDHRIHSDHSTSPSYSHKSSRDYRYSMDQMDHRAAGSGSPRSLDRSPYGSNPLG  
HRSPEHSDKSTPEHTWSRKT"  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"

BASE COUNT 2446 a 1223 c 1520 g 1683 t

ORIGIN

alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 ltleuProaspPProaspPlysProGlnAlaLysGlnLeuGlnth 17  
ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 4129  
4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 4129

17 rargAlaaspyrLeuileLysleuLeuasnLysaspleuAlaargLysg 34  
|||||  
4130 CCGTGCAGACTACCTCATTAATTAAGTAAGACCTTGCAGAAAG 4179

34 luAlaGlnarGleuAlaGlyAla 41  
|||||  
4180 AAGCACAAAGGCTTGCTGCTGCA 4202

seq\_name: gb\_pat:AS8684

seq\_documentation\_block:  
LOCUS AS8684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION AS8684  
VERSION AS8684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN



## BIRDS

JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 source Location/Qualifiers  
 1. 153  
 /db\_xref="taxon:32644"

BASE COUNT 58 a 40 c 31 g 24 t  
 ORIGIN

alignment\_scores:  
 Quality: 34.00 Length: 34  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x A58684 ..

Align seg 1/1 to: A58684 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 52 AAGAAACCCGAGCAAGCAAGCTGACAGCCGTCGAGACTACCTCATTTAA 101  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 102 ATTACTGAATTAAGACCTTGCAAGAAAGCAAGCAAGCTGCTGTG 151  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 41 1a 41  
 ||  
 152 CA 153

seq\_name: gb\_pat:A58685

seq\_documentation\_block:

LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 4 from Patent WO9639505.  
 ACCESSION A58685  
 VERSION A58685.1 GI:3714248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 AUTHORS Griffiths R. and Tlwardt B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 source Location/Qualifiers  
 1. 153  
 /db\_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t  
 ORIGIN

alignment\_scores:  
 Quality: 34.00 Length: 34  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x A58685 ..

Align seg 1/1 to: A58685 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 52 AAGAAACCCGAGCAAGCAAGCTGACAGCCGTCGAGACTACCTCATTTAA 101  
 ||||||||||||||||||||||||||||||||||||||||||||||||

24 sLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 41  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 102 ATTACTGAATTAAGACCTTGCAAGAAAGCAAGCAAGCTGCTGTG 151  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 41 1a 41  
 ||  
 152 CA 153

seq\_name: gb\_pat:A58686

seq\_documentation\_block:

LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 5 from Patent WO9639505.  
 ACCESSION A58686  
 VERSION A58686.1 GI:3714249  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 AUTHORS Griffiths R. and Tlwardt B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 source Location/Qualifiers  
 1. 153  
 /db\_xref="taxon:32644"

BASE COUNT 60 a 34 c 31 g 28 t  
 ORIGIN

alignment\_scores:  
 Quality: 27.00 Length: 27  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x A58686 ..

Align seg 1/1 to: A58686 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 52 AAGAAACCCGAGCAAGCAAGCTGACAGCCGTCGAGACTACCTCATTTAA 101  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 24 sLeuLeuAsnLysAspLeuAlaArgLysGln 34  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 102 ATTACTGAATTAAGACCTTGCAAGAAAGAA 132

seq\_name: gb\_ro:MUSCHD1X

seq\_documentation\_block:

LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
 DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
 ACCESSION L10410 X66028  
 VERSION L10410.1 GI:455014  
 KEYWORDS DNA binding protein.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 5349)  
 AUTHORS Delmas V., Stokes D.G. and Perry R.P.  
 TITLE A mammalian DNA-binding protein that contains a chromodomain and an SMC2/SMC4-like helix-loop domain  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
 MEDLINE 93211972  
 REFERENCE 2 (bases 1 to 5349)  
 AUTHORS Perry R.P.

TITLE Direct Submission  
JOURNAL Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES  
Location/Qualifiers

source  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDERSVRNGSGESSQSDGDCGSASGSGSSGSSSDGSS  
SOGSSDSDSGSGSSESSEDTRERKNVQAKPPKVDGAEPKWSPSILAQRASAML  
RKPOAOQOORPASNSGSESDSSSESDSSGAKRKHNDEDMOMSGSPOLG  
SDSESEERDSKSCDTESDYEPKKNVSRKPNRKSNGKKILGOKKROIDSDE  
DDEDYDNKRSRRQATVNSYKKEDEMKTSDDLLEYCGEDVPQDEEFTTIRM  
DCRVGRKATGATTTTAYEADGDNPAGPERKKEPGDIOYLTKMGWSHHTWETEE  
TLKQONVGMKKLDYKRRKDOETKRLMNASPEDEYVNCQDELTDLHKQIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSMEDALISKFOCIDYFSRNSKTIAPFK  
DCKVLKORPREVALKOPSYIGHEGLRDYQLNGMLAHSMCKGNSCIIADMG  
GKTIOITISFLNLYFHEHQLYGPFLVPLSTLSMORLETOMASOMNAVYIGDINSR  
NMIRHEMHQOTKRLKNILLTYEILLKOKARFAGFMASOMNAVYIGDINSR  
KTLIDFSNHRLLITGTPLONSLKEMLSLIMPEKFSMWEDPEERHKGREGYA  
SLHKELEPFLLRVAKDVEKSLPAKVEQILRHEMSALOKQYKWLITRYKALSGSK  
GSTGFLIMMELKCCNHCYLKPPDNNEFNRKQALOHILRSSGKLLDLKILRL  
BERGNRVLIIFSOMVRMLDLAEYLKRYOPFORLDGSLERKQALDHFNESSEDE  
CELSTFRAGGGINLASADTVIIFDSMNPNQDLQOARAHRIQOKOVNIYLVTKG  
SVEEDILERRAKKWLIDLVIORMDPTGKTIVHTGSASSSPKPKKEELSLKGAEE  
ELFKPEEBOEPQMDIDEITKRAETHENEGPLSVGDELLSOFKVFANFSMDEDDI  
ELEPERKSNMEELIPEOORRLEEEROKLELITMLPRMNCAKQISFNSSEGRS  
RSRRYSGSDSDSISERRKPKRGRPRPIPRENIKGSDAEIRRFISKYKFGEPERL  
DAIADAEIVKSEETDLRLGLVHNGCVKALKDSSGTERAGRLGKVGKGPFRISG  
VOVNAKLIYAHDELIRPLKSIIPDPEERKQYIIPCHTKAHEFIDMGKEDSNLLIG  
IYEVGSGWEMIKMDPOLSLTHKLLPDDPKKPOAKLOTRADYLIKLSRLAKREA  
ORLCAGSGSKRRKTRAKSKAMKSIKVEEIKSDSPLESEKDEDDKLDNARESK  
DRSKSVSDAIVHITASGEPIYIAESEBELDKTFISICKERMRPVKAKLOLDPEK  
GLSERQLEHTROCLIKIGDAITTECLKEYSNEQIKQWKNLIMIVSKATEADRLH  
KLYKAIRKROESQONSQNSNVATTHVIRNPMERLKEKNTNHDSSRDSYSDARLS  
QYHDKRHOQDSYKKSRSRRKPSFSGNSKXDRHMDHYRODSRYYSRPREKRLDD  
HRSRHSRLSGGLKDRCHSDHRSRSHSDHMSDHRSSSEHTHKKSSROYRILSDMQLD  
HRAASSGPRPLDORSPYGSRSPEHSHRSTPETHWSRKRT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x MUSCHDIX ..

Align seg 1/1 to: MUSCHDIX from: 1 to: 5349

1 11leuProaspasProaspIlyslsProglAlalatsglnleuGlnth 17  
|||||  
4026 ATTCTTCCCATGATCTGTGATATAAAACCAACGCAAAACAGTTACAGAC 4075  
17 TArgAlaAspTyrleuIleuIlyslsleu 26  
|||||  
4076 CCGTGCAGACTACCTCATCAACTACTT 4103

seq\_name: gp\_pr:AF006513  
seq\_documentation\_block:  
LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997

DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 5947)  
REFERENCE  
Woodage,T., Basrai,M.A., Baxevanis,A.D., Hieter,P. and Collins,F.S.  
Characterization of the CHD family of proteins  
Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE  
97470991  
2 (bases 1 to 5947)  
AUTHORS  
Woodage,T.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
Location/Qualifiers  
1..5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WI-5811"  
1..5947  
/gene="CHD1"  
164..5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db\_xref="GI:2645429"  
/translation="MNGHSDERSVRNGSGESSQSDGDCGSASGSGSSGSSSDGSS  
SOGSSDSDSGSGSSESSEDTRERKNVQAKPPKVDGAEPKWSPSILAQRASAML  
RKPOAOQOORPASNSGSESDSSSESDSSGAKRKHNDEDMOMSGSPOLG  
SDSESEERDSKSCDTESDYEPKKNVSRKPNRKSNGKKILGOKKROIDSSE  
DDEDYDNKRSRRQATVNSYKKEDEMKTSDDLLEYCGEDVPQDEEFTTIRM  
DCRVGRKATGATTTTAYEADGDNPAGPERKKEPGDIOYLTKMGWSHHTWETEE  
TLKQONVGMKKLDYKRRKDOETKRLMNASPEDEYVNCQDELTDLHKQIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSMEDALISKFOCIDYFSRNSKTIAPFK  
DCKVLKORPREVALKOPSYIGHEGLRDYQLNGMLAHSMCKGNSCIIADMG  
GKTIOITISFLNLYFHEHQLYGPFLVPLSTLSMORLETOMASOMNAVYIGDINSR  
NMIRHEMHQOTKRLKNILLTYEILLKOKARFAGFMASOMNAVYIGDINSR  
KTLIDFSNHRLLITGTPLONSLKEMLSLIMPEKFSMWEDPEERHKGREGYA  
SLHKELEPFLLRVAKDVEKSLPAKVEQILRHEMSALOKQYKWLITRYKALSGSK  
GSTGFLIMMELKCCNHCYLKPPDNNEFNRKQALOHILRSSGKLLDLKILRL  
BERGNRVLIIFSOMVRMLDLAEYLKRYOPFORLDGSLERKQALDHFNESSEDE  
CELSTFRAGGGINLASADTVIIFDSMNPNQDLQOARAHRIQOKOVNIYLVTKG  
SVEEDILERRAKKWLIDLVIORMDPTGKTIVHTGSASSSPKPKKEELSLKGAEE  
ELFKPEEBOEPQMDIDEITKRAETHENEGPLSVGDELLSOFKVFANFSMDEDDI  
ELEPERKSNMEELIPEOORRLEEEROKLELITMLPRMNCAKQISFNSSEGRS  
RSRRYSGSDSDSISERRKPKRGRPRPIPRENIKGSDAEIRRFISKYKFGEPERL  
DAIADAEIVKSEETDLRLGLVHNGCVKALKDSSGTERAGRLGKVGKGPFRISG  
VOVNAKLIYAHDELIRPLKSIIPDPEERKQYIIPCHTKAHEFIDMGKEDSNLLIG  
IYEVGSGWEMIKMDPOLSLTHKLLPDDPKKPOAKLOTRADYLIKLSRLAKREA  
ORLCAGSGSKRRKTRAKSKAMKSIKVEEIKSDSPLESEKDEDDKLDNARESK  
DRSKSVSDAIVHITASGEPIYIAESEBELDKTFISICKERMRPVKAKLOLDPEK  
GLSERQLEHTROCLIKIGDAITTECLKEYSNEQIKQWKNLIMIVSKATEADRLH  
KLYKAIRKROESQONSQNSNVATTHVIRNPMERLKEKNTNHDSSRDSYSDARLS  
QYHDKRHOQDSYKKSRSRRKPSFSGNSKXDRHMDHYRODSRYYSRPREKRLDD  
HRSRHSRLSGGLKDRCHSDHRSRSHSDHMSDHRSSSEHTHKKSSROYRILSDMQLD  
HRAASSGPRPLDORSPYGSRSPEHSHRSTPETHWSRKRT"

BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

```
US-08-973-363-8 x AF006513 ..
Align seg 1/1 to: AF006513 from: 1 to: 5947

1 HleuProaBspAProaBspAlysgProGlnAlaLysGlnLeuGlnTh 17
|||||
4025 ATCTCTCCAGATGATCCGATTAACCAACCAAGCAAGCAAGTTCGAGAC 4074
17 TARGAlAsPtyrLeuIleLysLeuLeu 26
|||||
4075 CCGTGCAGACTACTCATCAAAATTACTT 4102

seq_name: gb_pr:AC092372
seq_documentation_block:
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
ACCESSION AC092372
VERSION AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 101220)
DOE Joint Genome Institute.
REFERENCE
AUTHORS 2 (bases 1 to 101220)
Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS 3 (bases 1 to 101220)
Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
COMMENT
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-58M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x AC092372/rev ..
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 HleuProaBspAProaBspAlysgProGlnAlaLysGlnLeuGlnTh 17
|||||
26932 ATCTCTCCAGATGATCCGATTAACCAACCAAGCAAGCAAGTTCGAGAC 26883
17 TARGAlAsPtyrLeuIleLysLeuLeu 26

US-08-973-363-8 x AF006513 ..
Align seg 1/1 to: AF006513 from: 1 to: 5947

1 HleuProaBspAProaBspAlysgProGlnAlaLysGlnLeuGlnTh 17
|||||
4025 ATCTCTCCAGATGATCCGATTAACCAACCAAGCAAGCAAGTTCGAGAC 4074
17 TARGAlAsPtyrLeuIleLysLeuLeu 26
|||||
4075 CCGTGCAGACTACTCATCAAAATTACTT 4102

seq_name: gb_pr:AC012624
seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 134365)
DOE Joint Genome Institute.
REFERENCE
AUTHORS 2 (bases 1 to 134365)
Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE
AUTHORS 3 (bases 1 to 134365)
Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
COMMENT
FEATURES
source
1..134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365

1 HleuProaBspAProaBspAlysgProGlnAlaLysGlnLeuGlnTh 17
|||||
118247 ATCTCTCCAGATGATCCGATTAACCAACCAAGCAAGCAAGTTCGAGAC 118296
17 TARGAlAsPtyrLeuIleLysLeuLeu 26
|||||
118297 CCGTGCAGACTACTCATCAAAATTACTT 118324

seq_name: gb_htg:AC021449
seq_documentation_block:
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG; PHASE1; HTGS_DRAFT.
SOURCE human.
```

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 143079)  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 TITLE Homo sapiens, clone RP11-58M12  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 143079)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,  
 Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,  
 Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 DeArillano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,  
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,  
 Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K.,  
 MacDonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,  
 McPeeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
 Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,  
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
 Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:1407963.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L5154  
 Center clone name: 58\_M.12  
 Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 134743 bases at least Q40  
 Consensus quality: 139227 bases at least Q30  
 Consensus quality: 140814 bases at least Q20  
 Insert size: 144000; agarose-fp  
 Insert size: 142179; sum-of-coverage  
 Quality coverage: 4.6 in Q20 bases; agarose-fp  
 Quality coverage: 4.7 in Q20 bases; sum-of-coverage  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 10 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 38820: contig of 38820 bp in length  
 \* 38821 38920: gap of 100 bp  
 \* 38921 40411: contig of 1491 bp in length  
 \* 40412 40511: gap of 100 bp  
 \* 40512 43379: contig of 2768 bp in length  
 \* 43280 43379: gap of 100 bp  
 \* 43380 46905: contig of 3526 bp in length  
 \* 46906 47005: gap of 100 bp  
 \* 47006 51830: contig of 4825 bp in length  
 \* 51831 51930: gap of 100 bp  
 \* 51931 62619: contig of 10689 bp in length  
 \* 62620 62719: gap of 100 bp  
 \* 62720 75408: contig of 12689 bp in length

FEATURES  
 source  
 1. 143079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-58M12"  
 /clone\_11b="RP11 Human Male BAC"  
 1. 38820  
 /note="assembly-fragment"  
 clone\_end:sp6  
 vector\_side:left"  
 38921..40411  
 /note="assembly-fragment"  
 40512..43279  
 /note="assembly-fragment"  
 43380..46905  
 /note="assembly-fragment"  
 47006..51830  
 /note="assembly-fragment"  
 51931..62619  
 /note="assembly-fragment"  
 62720..75408  
 /note="assembly-fragment"  
 75509..92516  
 /note="assembly-fragment"  
 92617..106409  
 /note="assembly-fragment"  
 106510..143079  
 /note="assembly-fragment"  
 clone\_end:r7  
 vector\_side:right"  
 BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-8 x AC021449 ..  
 Align seq 1/1 to: AC021449 from: 1 to: 143079  
 1 11leupProAspPpProAspLysLysProGlnAlaLysGlnLeuGlnTH 17  
 116147 ATTCTTCACATGATGCCGATTAACCAACCAACAGATTGCAGAC 116196  
 17 rArGaLaSPtYrLeuLeuLysLeu 26  
 116197 CCGTGCAGACTACTCATCAATTACTT 116224  
 seq\_name: gp\_htg:AC008531  
 seq\_documentation\_block:  
 LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
 7 ordered pieces.  
 ACCESSION AC008531  
 VERSION AC008531.3 GI:12830078  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 145659)  
 AUTHORS DOE Joint Genome Institute.



```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5962 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 67981: gap of unknown length
* 67982 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117511: contig of 6605 bp in length
* 117512 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.

FEATURES
    source          1..193446
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="Rp11-36012"
                    /clone_lib="RPCT human BAC library 11"
BASE COUNT      57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-08-973-363-8 x AC091946 ..
Align seg 1/1 to: AC091946 from: 1 to: 193446
1 11leuProAspAspProAspLysProGlnAlaLysGlnLeuGln 17
|||||
87658 ATTCTTCCAGATGATCCCGATAAAAACCAACGCAAAACAGTTGCAGAC 87707
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
87708 CCGTGCAGACTACCTCATCAATTAATTACTT 87735
seq_name: gb_pr:AC026778
```

```
seq_documentation_block:
LOCUS      AC026778      195433 bp      DNA      linear      PRI 01-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.
ACCESSION  AC026778
VERSION    AC026778.4      GI:14277282
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 195433)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE Direct Submission
            JOURNAL Unpublished
            AUTHORS 2 (bases 1 to 195433)
            TITLE Direct Submission
            JOURNAL DOE Joint Genome Institute.
            AUTHORS 3 (bases 1 to 195433)
            TITLE Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            JOURNAL Direct Submission
            TITLE Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            JOURNAL On Jun 1, 2001 this sequence version replaced gi:13677045.
            COMMENT Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            WI-13675 G23101
            SHGC-58345 G38487
            SHGC-103595 G57841.

FEATURES
    source          1..195433
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTC-428111"
BASE COUNT      62762 a 37302 c 37040 g 58329 t
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-08-973-363-8 x AC026778/rev ..
Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433
1 11leuProAspAspProAspLysProGlnAlaLysGlnLeuGln 17
|||||
29732 ATTCTTCCAGATGATCCCGATAAAAACCAACGCAAAACAGTTGCAGAC 29683
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
29682 CCGTGCAGACTACCTCATCAATTAATTACTT 29655
seq_name: gb_pr:AC022121
seq_documentation_block:
LOCUS      AC022121      219258 bp      DNA      linear      PRI 30-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION  AC022121
VERSION    AC022121.6      GI:15375145
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
```



**THIS PAGE BLANK (USPTO)**





1188 TCGTGCAGATTACCTCATTAATTACTGAATAAAGACCTTGCAAGAAAG 1237

34 luAaIaInArGLeuAlaGlyAla 41  
|||||  
1238 AAGCAGACGAGACTGCTGTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

AC AA142751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT CDS 228..5390

FT /tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 1; Fig 5; 76pp; English.

CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AA142751) is suggested to initiate

CC female development in birds. The sequence of CDH-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AA142755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC genome (see also AA142756-57). It is located on an autosome or Z

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-rare

CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

SO

alignment\_scores:

Quality: 41.00 Length: 41

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lIleuPProAspPProAspLysLysProGlnAlaLysGlnLeuGlnH 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGCAGCTACAGAC 4129

17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34

|||||  
4130 CCGTGCAGACTACCTCATTAATTACTGAATAAAGACCTTGCAAGAAAG 4179

34 luAaIaInArGLeuAlaGlyAla 41  
|||||

4180 AAGCAGAAAGCCTGCTGTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq\_documentation\_block:

ID AA142757 standard; DNA; 153 BP.

AC AA142757;

DT 12-MAR-1997 (first entry)

DE Chick CHD-1A gene fragment.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

and are ignored in the translated amino

acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

P-PSDB: AAW08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757).

CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)

CC and the great tit CHD-W gene (see also AA142759). Translated amino

CC acid sequences of this region are provided in AAW08146-49. The

CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes

CC determine sex in birds and can be used to identify the sex of an

XX embryo, foetus etc. and to manipulate the sex of progeny.

SO Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 34.00 Length: 34

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x AA142757 ..

Align seg 1/1 to: AA142757 from: 1 to: 153

```

      8 LysLysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleLeu 24
      52 AAGAAACCCAGCAGCAAGACGCTACAGACCCGTCAGACTTACTCATTTAA 101
      24 steuleuasnlysaAspleuAlaArgLysGlnAlaGlnArgLeuAlaGlnYA 41
      102 ATTACTGAATAAAGACCTTGCAAGAAAGACACAGAAAGCCTTCTGCTG 151
      41 la 41
      152 CA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758

seq_documentation_block:
ID   AA142758 standard; DNA; 153 BP.
AC   AA142758;
XX
XX
DT   12-MAR-1997 (first entry)
XX
DE   Chick CHD-W gene fragment.
XX
XX   Bird; sex determination: chromodomain-Helicase-DNA binding 1;
KM   CHD-1A; CHD-W; W chromosome; ss.
XX
OS   Gallus sp.
XX
XX   Key Location/Qualifiers
FH   misc_difference 52..81
FT   /tag= a
FT   /note= "bases 52-81 are a repeat of bases 22-51
FT   and are ignored in the translated amino
FT   acid sequence given in Fig 3"
XX
XX   WO9639505-A1.
XX
XX   12-DEC-1996.
XX
XX   05-JUN-1996; 96WO-GB01341.
XX
XX   06-JUN-1995; 95GB-0011439.
XX
XX   (ISIS-) ISIS INNOVATION LTD.
XX
XX   Griffiths R, Tiwari B;
XX
XX   WPI: 1997-043127/04.
XX
XX   P-PSDB: AAW08148.
XX
XX   Claim 8; Fig 3; 76pp; English.
XX
XX   Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC   to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
CC   chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC   and the great tit CHD-W gene (AA142759). Translated amino acid
CC   sequences of this region are provided in AAW08146-49. The CHD-1A
CC   (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC   sex in birds and can be used to identify the sex of an embryo.
CC   foetus etc. and to manipulate the sex of progeny.
XX
XX   Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

      8 LysLysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleLeu 24
      52 AAGAAACCCAGCAGCTAAGCATACGACCCGTCAGACTTACTCATTTAA 101
      24 steuleuasnlysaAspleuAlaArgLysGlnAlaGlnArgLeuAlaGlnYA 41
      102 ATTACTGAATAAAGACCTTGCAAGAAAGACACAGAAAGCCTTCTGCTG 151
      41 la 41
      152 CA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq_documentation_block:
ID   AA142759 standard; DNA; 153 BP.
AC   AA142759;
XX
XX
DT   12-MAR-1997 (first entry)
XX
DE   Great tit CHD-W gene fragment.
XX
XX   Bird; sex determination: chromodomain-Helicase-DNA binding 1;
KM   CHD-1A; CHD-W; W chromosome; ss.
XX
XX   Parus major.
XX
XX   Key Location/Qualifiers
FH   misc_difference 52..81
FT   /tag= a
FT   /note= "bases 52-81 are a repeat of bases 22-51
FT   and are ignored in the translated amino
FT   acid sequence given in Fig 3"
XX
XX   WO9639505-A1.
XX
XX   12-DEC-1996.
XX
XX   05-JUN-1996; 96WO-GB01341.
XX
XX   06-JUN-1995; 95GB-0011439.
XX
XX   (ISIS-) ISIS INNOVATION LTD.
XX
XX   Griffiths R, Tiwari B;
XX
XX   WPI: 1997-043127/04.
XX
XX   P-PSDB: AAW08149.
XX
XX   Claim 8; Fig 3; 76pp; English.
XX
XX   Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
CC   to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
CC   chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC   and the great tit CHD-W gene (AA142759). Translated amino acid
CC   sequences of this region are provided in AAW08146-49. The CHD-1A
CC   (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC   sex in birds and can be used to identify the sex of an embryo.
CC   foetus etc. and to manipulate the sex of progeny.
XX
XX   Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment_scores:

```



CC progeny.  
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
SQ

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

8 lysysproglAlalysglInleugInthRargAlaAspyrleuileally 24  
|||||  
52 AAAAAACCCAGCAGCAGCAGCTTACAGACCCGTCGACACTCATCACA 101  
|||||

24 sleuleu 26  
|||||  
102 ACTACTT 108

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:  
ID ABL06443 standard; cDNA; 6240 BP.  
XX  
AC ABL06443;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
PI  
DR WPI: 2001-656860/75.  
DR P-PSDB: ABB62340.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins  
XX (AB57737-AB572072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;  
XX

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

9 lysproglAlalysglInleugInthRargAla 19  
|||||  
4205 AACCAGCCAGCAGCAGCTGACAGCGGTC 4237  
|||||

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:  
ID ABL06442 standard; cDNA; 9933 BP.  
XX  
AC ABL06442;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
PI  
DR WPI: 2001-656860/75.  
DR P-PSDB: ABB62339.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins  
XX (AB57737-AB572072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0

Percent similarity: 100.000 Percent identity: 100.000

alignment block:  
US-08-973-363-8 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
6745 AGCCCCAGCCAGCAGCTGCAGAGCGCTGCC 6777

seq\_name: /SIDS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882

seq\_documentation\_block:

ID AAK88882 standard; cDNA: 421 BP.

XX AAK88882;

DT 05-NOV-2001 (first entry)

DE Human digestive system antigen coding sequence SEQ ID NO: 1198.

KW Human; digestive system antigen; gene therapy; cancer; appendicitis;

KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;

KW digestive system disorder; Meckel's diverticulum; ss.

XX Homo sapiens.

PN WO200155314-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01324.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235483.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251969.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-502630/55.  
DR P-PSDB; AAM93109.  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosis, treating, preventing and/or prognostic disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Claim 1; SEQ ID NO 1198; 986bp; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders.  
CC including cancer Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
XX  
Alignment\_scores:                   Quality: 10.00                   Length: 10  
                                  Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000  
XX  
alignment\_block:  
US-08-973-363-8 x AAK88882 ..  
XX  
Align seg 1/1 to: AAK88882 from: 1 to: 421  
XX  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
          |||||  
146 AAGCAGCTACAGACCCGACGCGATTACTTG 175  
XX  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseg-emb1/NA2001A.DAT:AA157603  
seq\_documentation\_block:  
ID AA157603 standard; cDNA: 421 BP.  
XX  
AC AA157603;  
XX  
AT 19-OCT-2001 (first entry)  
XX  
DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.

XX  
KW Human: colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
XX Homo sapiens.  
OS WO20015350-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01350.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0186874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAR-2000; 2000US-0205315.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225478.  
PR 14-AUG-2000; 2000US-0225479.  
PR 14-AUG-2000; 2000US-0225480.  
PR 14-AUG-2000; 2000US-0225481.  
PR 14-AUG-2000; 2000US-0225482.  
PR 14-AUG-2000; 2000US-0225483.  
PR 14-AUG-2000; 2000US-0225484.  
PR 14-AUG-2000; 2000US-0225485.  
PR 14-AUG-2000; 2000US-0225486.  
PR 14-AUG-2000; 2000US-0225487.  
PR 14-AUG-2000; 2000US-0225488.  
PR 14-AUG-2000; 2000US-0225489.  
PR 14-AUG-2000; 2000US-0225490.  
PR 14-AUG-2000; 2000US-0225491.  
PR 14-AUG-2000; 2000US-0225492.  
PR 14-AUG-2000; 2000US-0225493.  
PR 14-AUG-2000; 2000US-0225494.  
PR 14-AUG-2000; 2000US-0225495.  
PR 14-AUG-2000; 2000US-0225496.  
PR 14-AUG-2000; 2000US-0225497.  
PR 14-AUG-2000; 2000US-0225498.  
PR 14-AUG-2000; 2000US-0225499.  
PR 14-AUG-2000; 2000US-0225500.  
PR 14-AUG-2000; 2000US-0225501.  
PR 14-AUG-2000; 2000US-0225502.  
PR 14-AUG-2000; 2000US-0225503.  
PR 14-AUG-2000; 2000US-0225504.  
PR 14-AUG-2000; 2000US-0225505.  
PR 14-AUG-2000; 2000US-0225506.  
PR 14-AUG-2000; 2000US-0225507.  
PR 14-AUG-2000; 2000US-0225508.  
PR 14-AUG-2000; 2000US-0225509.  
PR 14-AUG-2000; 2000US-0225510.  
PR 14-AUG-2000; 2000US-0225511.  
PR 14-AUG-2000; 2000US-0225512.  
PR 14-AUG-2000; 2000US-0225513.  
PR 14-AUG-2000; 2000US-0225514.  
PR 14-AUG-2000; 2000US-0225515.  
PR 14-AUG-2000; 2000US-0225516.  
PR 14-AUG-2000; 2000US-0225517.  
PR 14-AUG-2000; 2000US-0225518.  
PR 14-AUG-2000; 2000US-0225519.  
PR 14-AUG-2000; 2000US-0225520.  
PR 14-AUG-2000; 2000US-0225521.  
PR 14-AUG-2000; 2000US-0225522.  
PR 14-AUG-2000; 2000US-0225523.  
PR 14-AUG-2000; 2000US-0225524.  
PR 14-AUG-2000; 2000US-0225525.  
PR 14-AUG-2000; 2000US-0225526.  
PR 14-AUG-2000; 2000US-0225527.  
PR 14-AUG-2000; 2000US-0225528.  
PR 14-AUG-2000; 2000US-0225529.  
PR 14-AUG-2000; 2000US-0225530.  
PR 14-AUG-2000; 2000US-0225531.  
PR 14-AUG-2000; 2000US-0225532.  
PR 14-AUG-2000; 2000US-0225533.  
PR 14-AUG-2000; 2000US-0225534.  
PR 14-AUG-2000; 2000US-0225535.  
PR 14-AUG-2000; 2000US-0225536.  
PR 14-AUG-2000; 2000US-0225537.  
PR 14-AUG-2000; 2000US-0225538.  
PR 14-AUG-2000; 2000US-0225539.  
PR 14-AUG-2000; 2000US-0225540.  
PR 14-AUG-2000; 2000US-0225541.  
PR 14-AUG-2000; 2000US-0225542.  
PR 14-AUG-2000; 2000US-0225543.  
PR 14-AUG-2000; 2000US-0225544.  
PR 14-AUG-2000; 2000US-0225545.  
PR 14-AUG-2000; 2000US-0225546.  
PR 14-AUG-2000; 2000US-0225547.  
PR 14-AUG-2000; 2000US-0225548.  
PR 14-AUG-2000; 2000US-0225549.  
PR 14-AUG-2000; 2000US-0225550.  
PR 14-AUG-2000; 2000US-0225551.  
PR 14-AUG-2000; 2000US-0225552.  
PR 14-AUG-2000; 2000US-0225553.  
PR 14-AUG-2000; 2000US-0225554.  
PR 14-AUG-2000; 2000US-0225555.  
PR 14-AUG-2000; 2000US-0225556.  
PR 14-AUG-2000; 2000US-0225557.  
PR 14-AUG-2000; 2000US-0225558.  
PR 14-AUG-2000; 2000US-0225559.  
PR 14-AUG-2000; 2000US-0225560.  
PR 14-AUG-2000; 2000US-0225561.  
PR 14-AUG-2000; 2000US-0225562.  
PR 14-AUG-2000; 2000US-0225563.  
PR 14-AUG-2000; 2000US-0225564.  
PR 14-AUG-2000; 2000US-0225565.  
PR 14-AUG-2000; 2000US-0225566.  
PR 14-AUG-2000; 2000US-0225567.  
PR 14-AUG-2000; 2000US-0225568.  
PR 14-AUG-2000; 2000US-0225569.  
PR 14-AUG-2000; 2000US-0225570.  
PR 14-AUG-2000; 2000US-0225571.  
PR 14-AUG-2000; 2000US-0225572.  
PR 14-AUG-2000; 2000US-0225573.  
PR 14-AUG-2000; 2000US-0225574.  
PR 14-AUG-2000; 2000US-0225575.  
PR 14-AUG-2000; 2000US-0225576.  
PR 14-AUG-2000; 2000US-0225577.  
PR 14-AUG-2000; 2000US-0225578.  
PR 14-AUG-2000; 2000US-0225579.  
PR 14-AUG-2000; 2000US-0225580.  
PR 14-AUG-2000; 2000US-0225581.  
PR 14-AUG-2000; 2000US-0225582.  
PR 14-AUG-2000; 2000US-0225583.  
PR 14-AUG-2000; 2000US-0225584.  
PR 14-AUG-2000; 2000US-0225585.  
PR 14-AUG-2000; 2000US-0225586.  
PR 14-AUG-2000; 2000US-0225587.  
PR 14-AUG-2000; 2000US-0225588.  
PR 14-AUG-2000; 2000US-0225589.  
PR 14-AUG-2000; 2000US-0225590.  
PR 14-AUG-2000; 2000US-0225591.  
PR 14-AUG-2000; 2000US-0225592.  
PR 14-AUG-2000; 2000US-0225593.  
PR 14-AUG-2000; 2000US-0225594.  
PR 14-AUG-2000; 2000US-0225595.  
PR 14-AUG-2000; 2000US-0225596.  
PR 14-AUG-2000; 2000US-0225597.  
PR 14-AUG-2000; 2000US-0225598.  
PR 14-AUG-2000; 2000US-0225599.  
PR 14-AUG-2000; 2000US-0225600.  
PR 14-AUG-2000; 2000US-0225601.  
PR 14-AUG-2000; 2000US-0225602.  
PR 14-AUG-2000; 2000US-0225603.  
PR 14-AUG-2000; 2000US-0225604.  
PR 14-AUG-2000; 2000US-0225605.  
PR 14-AUG-2000; 2000US-0225606.  
PR 14-AUG-2000; 2000US-0225607.  
PR 14-AUG-2000; 2000US-0225608.  
PR 14-AUG-2000; 2000US-0225609.  
PR 14-AUG-2000; 2000US-0225610.  
PR 14-AUG-2000; 2000US-0225611.  
PR 14-AUG-2000; 2000US-0225612.  
PR 14-AUG-2000; 2000US-0225613.  
PR 14-AUG-2000; 2000US-0225614.  
PR 14-AUG-2000; 2000US-0225615.  
PR 14-AUG-2000; 2000US-0225616.  
PR 14-AUG-2000; 2000US-0225617.  
PR 14-AUG-2000; 2000US-0225618.  
PR 14-AUG-2000; 2000US-0225619.  
PR 14-AUG-2000; 2000US-0225620.  
PR 14-AUG-2000; 2000US-0225621.  
PR 14-AUG-2000; 2000US-0225622.  
PR 14-AUG-2000; 2000US-0225623.  
PR 14-AUG-2000; 2000US-0225624.  
PR 14-AUG-2000; 2000US-0225625.  
PR 14-AUG-2000; 2000US-0225626.  
PR 14-AUG-2000; 2000US-0225627.  
PR 14-AUG-2000; 2000US-0225628.  
PR 14-AUG-2000; 2000US-0225629.  
PR 14-AUG-2000; 2000US-0225630.  
PR 14-AUG-2000; 2000US-0225631.  
PR 14-AUG-2000; 2000US-0225632.  
PR 14-AUG-2000; 2000US-0225633.  
PR 14-AUG-2000; 2000US-0225634.  
PR 14-AUG-2000; 2000US-0225635.  
PR 14-AUG-2000; 2000US-0225636.  
PR 14-AUG-2000; 2000US-0225637.  
PR 14-AUG-2000; 2000US-0225638.  
PR 14-AUG-2000; 2000US-0225639.  
PR 14-AUG-2000; 2000US-0225640.  
PR 14-AUG-2000; 2000US-0225641.  
PR 14-AUG-2000; 2000US-0225642.  
PR 14-AUG-2000; 2000US-0225643.  
PR 14-AUG-2000; 2000US-0225644.  
PR 14-AUG-2000; 2000US-0225645.  
PR 14-AUG-2000; 2000US-0225646.  
PR 14-AUG-2000; 2000US-0225647.  
PR 14-AUG-2000; 2000US-0225648.  
PR 14-AUG-2000; 2000US-0225649.  
PR 14-AUG-2000; 2000US-0225650.  
PR 14-AUG-2000; 2000US-0225651.  
PR 14-AUG-2000; 2000US-0225652.  
PR 14-AUG-2000; 2000US-0225653.  
PR 14-AUG-2000; 2000US-0225654.  
PR 14-AUG-2000; 2000US-0225655.  
PR 14-AUG-2000; 2000US-0225656.  
PR 14-AUG-2000; 2000US-0225657.  
PR 14-AUG-2000; 2000US-0225658.  
PR 14-AUG-2000; 2000US-0225659.  
PR 14-AUG-2000; 2000US-0225660.  
PR 14-AUG-2000; 2000US-0225661.  
PR 14-AUG-2000; 2000US-0225662.  
PR 14-AUG-2000; 2000US-0225663.  
PR 14-AUG-2000; 2000US-0225664.  
PR 14-AUG-2000; 2000US-0225665.  
PR 14-AUG-2000; 2000US-0225666.  
PR 14-AUG-2000; 2000US-0225667.  
PR 14-AUG-2000; 2000US-0225668.  
PR 14-AUG-2000; 2000US-0225669.  
PR 14-AUG-2000; 2000US-0225670.  
PR 14-AUG-2000; 2000US-0225671.  
PR 14-AUG-2000; 2000US-0225672.  
PR 14-AUG-2000; 2000US-0225673.  
PR 14-AUG-2000; 2000US-0225674.  
PR 14-AUG-2000; 2000US-0225675.  
PR 14-AUG-2000; 2000US-0225676.  
PR 14-AUG-2000; 2000US-0225677.  
PR 14-AUG-2000; 2000US-0225678.  
PR 14-AUG-2000; 2000US-0225679.  
PR 14-AUG-2000; 2000US-0225680.  
PR 14-AUG-2000; 2000US-0225681.  
PR 14-AUG-2000; 2000US-0225682.  
PR 14-AUG-2000; 2000US-0225683.  
PR 14-AUG-2000; 2000US-0225684.  
PR 14-AUG-2000; 2000US-0225685.  
PR 14-AUG-2000; 2000US-0225686.  
PR 14-AUG-2000; 2000US-0225687.  
PR 14-AUG-2000; 2000US-0225688.  
PR 14-AUG-2000; 2000US-0225689.  
PR 14-AUG-2000; 2000US-0225690.  
PR 14-AUG-2000; 2000US-0225691.  
PR 14-AUG-2000; 2000US-0225692.  
PR 14-AUG-2000; 2000US-0225693.  
PR 14-AUG-2000; 2000US-0225694.  
PR 14-AUG-2000; 2000US-0225695.  
PR 14-AUG-2000; 2000US-0225696.  
PR 14-AUG-2000; 2000US-0225697.  
PR 14-AUG-2000; 2000US-0225698.  
PR 14-AUG-2000; 2000US-0225699.  
PR 14-AUG-2000; 2000US-0225700.  
PR 14-AUG-2000; 2000US-0225701.  
PR 14-AUG-2000; 2000US-0225702.  
PR 14-AUG-2000; 2000US-0225703.  
PR 14-AUG-2000; 2000US-0225704.  
PR 14-AUG-2000; 2000US-0225705.  
PR 14-AUG-2000; 2000US-0225706.  
PR 14-AUG-2000; 2000US-0225707.  
PR 14-AUG-2000; 2000US-0225708.  
PR 14-AUG-2000; 2000US-0225709.  
PR 14-AUG-2000; 2000US-0225710.  
PR 14-AUG-2000; 2000US-0225711.  
PR 14-AUG-2000; 2000US-0225712.  
PR 14-AUG-2000; 2000US-0225713.  
PR 14-AUG-2000; 2000US-0225714.  
PR 14-AUG-2000; 2000US-0225715.  
PR 14-AUG-2000; 2000US-0225716.  
PR 14-AUG-2000; 2000US-0225717.  
PR 14-AUG-2000; 2000US-0225718.  
PR 14-AUG-2000; 2000US-0225719.  
PR 14-AUG-2000; 2000US-0225720.  
PR 14-AUG-2000; 2000US-0225721.  
PR 14-AUG-2000; 2000US-0225722.  
PR 14-AUG-2000; 2000US-0225723.  
PR 14-AUG-2000; 2000US-0225724.  
PR 14-AUG-2000; 2000US-0225725.  
PR 14-AUG-2000; 2000US-0225726.  
PR 14-AUG-2000; 2000US-0225727.  
PR 14-AUG-2000; 2000US-0225728.  
PR 14-AUG-2000; 2000US-0225729.  
PR 14-AUG-2000; 2000US-0225730.  
PR 14-AUG-2000; 2000US-0225731.  
PR 14-AUG-2000; 2000US-0225732.  
PR 14-AUG-2000; 2000US-0225733.  
PR 14-AUG-2000; 2000US-0225734.  
PR 14-AUG-2000; 2000US-0225735.  
PR 14-AUG-2000; 2000US-0225736.  
PR 14-AUG-2000; 2000US-0225737.  
PR 14-AUG-2000; 2000US-0225738.  
PR 14-AUG-2000; 2000US-0225739.  
PR 14-AUG-2000; 2000US-0225740.  
PR 14-AUG-2000; 2000US-0225741.  
PR 14-AUG-2000; 2000US-0225742.  
PR 14-AUG-2000; 2000US-0225743.  
PR 14-AUG-2000; 2000US-0225744.  
PR 14-AUG-2000; 2000US-0225745.  
PR 14-AUG-2000; 2000US-0225746.  
PR 14-AUG-2000; 2000US-0225747.  
PR 14-AUG-2000; 2000US-0225748.  
PR 14-AUG-2000; 2000US-0225749.  
PR 14-AUG-2000; 2000US-0225750.  
PR 14-AUG-2000; 2000US-0225751.  
PR 14-AUG-2000; 2000US-0225752.  
PR 14-AUG-2000; 2000US-0225753.  
PR 14-AUG-2000; 2000US-0225754.  
PR 14-AUG-2000; 2000US-0225755.  
PR 14-AUG-2000; 2000US-0225756.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225760.  
PR 14-AUG-2000; 2000US-0225761.  
PR 14-AUG-2000; 2000US-0225762.  
PR 14-AUG-2000; 2000US-0225763.  
PR 14-AUG-2000; 2000US-0225764.  
PR 14-AUG-2000; 2000US-0225765.  
PR 14-AUG-2000; 2000US-0225766.  
PR 14-AUG-2000; 2000US-0225767.  
PR 14-AUG-2000; 2000US-0225768.  
PR 14-AUG-2000; 2000US-0225769.  
PR 14-AUG-2000; 2000US-0225770.  
PR 14-AUG-2000; 2000US-0225771.  
PR 14-AUG-2000; 2000US-0225772.  
PR 14-AUG-2000; 2000US-0225773.  
PR 14-AUG-2000; 2000US-0225774.  
PR 14-AUG-2000; 2000US-0225775.  
PR 14-AUG-2000; 2000US-0225776.  
PR 14-AUG-2000; 2000US-0225777.  
PR 14-AUG-2000; 2000US-0225778.  
PR 14-AUG-2000; 2000US-0225779.  
PR 14-AUG-2000; 2000US-0225780.  
PR 14-AUG-2000; 2000US-0225781.  
PR 14-AUG-2000; 2000US-0225782.  
PR 14-AUG-2000; 2000US-0225783.  
PR 14-AUG-2000; 2000US-0225784.  
PR 14-AUG-2000; 2000US-0225785.  
PR 14-AUG-2000; 2000US-0225786.  
PR 14-AUG-2000; 2000US-0225787.  
PR 14-AUG-2000; 2000US-0225788.  
PR 14-AUG-2000; 2000US-0225789.  
PR 14-AUG-2000; 2000US-0225790.  
PR 14-AUG-2000; 2000US-0225791.  
PR 14-AUG-2000; 2000US-0225792.  
PR 14-AUG-2000; 2000US-0225793.  
PR 14-AUG-2000; 2000US-0225794.  
PR 14-AUG-2000; 2000US-0225795.  
PR 14-AUG-2000; 2000US-0225796.  
PR 14-AUG-2000; 2000US-0225797.  
PR 14-AUG-2000; 2000US-0225798.  
PR 14-AUG-2000; 2000US-0225799.  
PR 14-AUG-2000; 2000US-0225800.  
PR 14-AUG-2000; 2000US-0225801.  
PR 14-AUG-2000; 2000US-0225802.  
PR 14-AUG-2000; 2000US-0225803.  
PR 14-AUG-2000; 2000US-0225804.  
PR 14-AUG-2000; 2000US-0225805.  
PR 14-AUG-2000; 2000US-0225806.  
PR 14-AUG-2000; 2000US-0225807.  
PR 14-AUG-2000; 2000US-0225808.  
PR 14-AUG-2000; 2000US-0225809.  
PR 14-AUG-2000; 2000US-0225810.  
PR 14-AUG-2000; 2000US-0225811.  
PR 14-AUG-2000; 2000US-0225812.  
PR 14-AUG-2000; 2000US-0225813.  
PR 14-AUG-2000; 2000US-0225814.  
PR 14-AUG-2000; 2000US-0225815.  
PR 14-AUG-2000; 2000US-0225816.  
PR 14-AUG-2000; 2000US-0225817.  
PR 14-AUG-2000; 2000US-0225818.  
PR 14-AUG-2000; 2000US-0225819.  
PR 14-AUG-2000; 2000US-0225820.  
PR





```
XX Human kinase poly(nucleotide(s) and recombinant products - useful
PT for identification of modulators of the enzyme, and treatment of
PT diseases associated with abnormal kinase expression
XX
XX Claim 1; Columns 29-30; 30pp; English.
XX
CC The invention provides polynucleotides which encode novel protein
CC kinase homologs expressed in various human cells and tissues. The
CC present sequence represents the clone 23789 cDNA fragment derived
CC from a human inflamed adenoid cDNA library. The cDNA encodes a
CC protein kinase which shows homology to the CKRPNK chicken tyr kinase.
CC Vectors and host cells can be used for recombinant production of the
CC protein kinase homolog. The recombinant proteins may be used to raise
CC antibodies for use as anti-kinase therapeutics. Oligonucleotides based
CC on the polynucleotide sequences, i.e. probes and antisense constructs,
CC the peptides and antibodies are claimed to be useful as tools for
CC studying signalling cascades in cells and proteins, and for identifying
CC inhibitors (drugs) to treat diseases and inflammatory conditions
CC associated with abnormal kinase expression. Diseases that are claimed
CC to be treatable include human x-linked agammaglobulinemia,
CC nonspherocytic hemolytic anaemia, artherosclerosis, cardiomyas,
CC diabetes, gliomas, restenosis, cholera-based septic shock, etc.
XX
SQ Sequence 252 BP; 63 A; 70 C; 72 G; 47 T; 0 other;

alignment_scores:
      Quality: 8.00      Length: 8
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x AAV33521 ..

Align seg 1/1 to: AAV33521 from: 1 to: 252

24 LysLeuLeuAsnLysAspLeuAla 31
|||||
100 AACCTGCTCAACAAAGACCTGCA 123

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA61903
seq_documentation_block:
ID ABA61903 standard; DNA: 544 BP.
XX
XX ABA61903:
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #10208.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PT
```

```
XX WPI; 2001-483447/52.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver.
XX
XX Claim 1; SEQ ID NO 10208; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WPI at ftp.wpi.int/pub/published_ptc_sequences.
XX
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
      Quality: 8.00      Length: 8
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x ABA61903 ..

Align seg 1/1 to: ABA61903 from: 1 to: 544

21 TyrLeuLeuLysLeuLeuAsnLys 28
|||||
243 TACCTGATTAAGCTTATTAATAATAG 266

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK10214
seq_documentation_block:
ID AAK10214 standard; DNA: 544 BP.
XX
XX AAK10214:
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon SEQ ID NO: 10205.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PT WPI; 2001-483446/52.
PT Single exon nucleic acid probes for analyzing gene expression in human
```

PT brains -  
XX  
PS Example 4; SEQ ID NO: 10205; 650bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system,  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-8 x AAK10214 ..  
  
Align seg 1/1 to: AAK10214 from: 1 to: 544  
  
21 TyrlleuilelysleuLeuAnlys 28  
|||||  
243 TACCTGATTAAGTTATTAATAAG 266  
  
seq\_name: /SIS1/gcgcdata/hold-geneseq/genesegn-emb1/NA2001A.DAT:AAK36112  
  
seq\_documentation\_block:  
ID AAK36112 standard; DNA; 544 BP.  
XX  
AC AAK36112:  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.  
XX  
KW Human: bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
WPI; 2001-488900/53.  
XX  
DR Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 10669; 658bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow

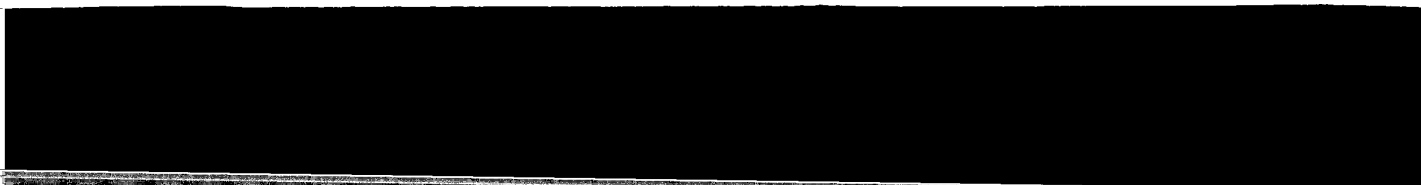
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-8 x AAK36112 ..  
  
Align seg 1/1 to: AAK36112 from: 1 to: 544  
  
21 TyrlleuilelysleuLeuAnlys 28  
|||||  
243 TACCTGATTAAGTTATTAATAAG 266

Mon Aug 5 11:52:00 2002

us-08-973-363-8.olip2n.rng

Page 11

---



**THIS PAGE BLANK (USP 10)**



DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 645)  
AUTHORS Hucke, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Hucke E  
Sanger Centre  
Hinxtun, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: L1E1d12.pic  
Sequencing primer: PIC  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
FEATURES  
Location/Qualifiers  
1..645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone\_lib="L1E1d12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"  
BASE COUNT 222 a 125 c 156 g 141 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 30.00 Length: 30  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AL644594 ..  
Align seg 1/1 to: AL644594 from: 1 to: 645  
1 TtleupProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnrth 17  
|||||  
456 ATTTCACAGATGATCCAGATAGAACCCCAAGCCCAAGCAGCTACAGAC 505  
17 fArgAlaAspTyrLeuIleLysLeuAsnLysAspLeu 30  
|||||  
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGATCTG 545  
seq\_name: gb\_est1:AL601246  
seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKF2P313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKF2P313J1040 5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann, S.

TITLE EST (Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sanger Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No SI sequence available.  
This clone (DKF2P313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
FEATURES  
Location/Qualifiers  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="DKF2P313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiI; Site\_2: SfiIb; cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AL601246 ..  
Align seg 1/1 to: AL601246 from: 1 to: 430  
1 TtleupProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnrth 17  
|||||  
218 ATTTCACAGATGATCCGATTAAGAACCAAGCAAGCAGTTGACAGAC 267  
17 fArgAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
268 CCGTGCAGACTACCTCATCAATTACTT 295  
seq\_name: gb\_est1:AI890775  
seq\_documentation\_block:  
LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wm95f11.x1 NCI CGAP Ut2 Homo sapiens cDNA clone IMAGE:2443725 3',  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; , mRNA sequence.  
ACCESSION AI890775  
VERSION AI890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmerit-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.



1 l1leuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
489 ATTCTTCCAGATGATCCTGATTAACCAAGCAAGCAAGTACAGAC 538  
17 fARGAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
539 CCGTGCAGACTACCTCATCAACTACTT 566

seq\_name: gb\_est1:BB461065

seq\_documentation\_block:

LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to l10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION BB461065  
VERSION BB461065.2 GI:16426612

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
'D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeeda,Y., Tanaka,T., Toyota,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

TITLE  
JOURNAL  
COMMENT On Jul 21, 2000 this sequence version replaced gi:9356558.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
'M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
'S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
(11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
'Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamano,I., Aizawa,  
'K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

FEATURES  
Source 1. 660  
Location/Qualifiers

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_1ib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site: 1: Saito, Site: 2: BamHI, cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGCGCGCAACTCGAGATTGTTTCTTTTATCCCTCCCTCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 211 a 130 c 168 g 151 t

ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x BB461065 ..  
Align seg 1/1 to: BB461065 from: 1 to: 660

1 l1leuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
481 ATTCTTCCAGATGATCCTGATTAACCAAGCAAGCAAGTACAGAC 530

17 fARGAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
531 CCGTGCAGACTACCTCATCAACTACTT 558

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.

ACCESSION AU125712  
VERSION AU125712.1 GI:10950428

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,  
'Y., Sugano,S., Isogai,T.)  
Unpublished (2000)

JOURNAL  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1533-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp



HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES  
Location/Qualifiers

1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_id="NT2RM4"  
/cell\_type="keratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SF13; mRNA from uninduced NT2 neuronal precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AU125712

Align seg 1/1 to: AU125712 from: 1 to: 866

1 lleleuproaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
450 ATTCTTCGAGATGATCCGATTAACACCAAGCAAGCAAGTTCGACAC 499  
17 rArgAlaasPtyrLeuIleLysLeuLeu 26  
|||||  
500 CCGTCGACGACCTACCTCATCAATTCCTT 527

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436080F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.

ACCESSION BE895133  
VERSION BE895133.1 GI:10358221

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1028)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: L1469753 row: h column: 16  
High quality sequence stop: 488.

FEATURES  
Location/Qualifiers  
1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_id="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; site\_1: NotI;

Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x BE895133

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 lleleuproaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
131 ATTCTTCGAGATGATCCGATTAACACCAAGCAAGCAAGTTCGACAC 180  
17 rArgAlaasPtyrLeuIleLysLeuLeu 26  
|||||  
181 CCGTCGACGACCTACCTCATCAATTCCTT 208

seq\_name: gb\_est1:AW96787

seq\_documentation\_block:

LOCUS AW96787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW96787  
VERSION AW96787.1 GI:8257021

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 337)

AUTHORS Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,  
Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,  
Goldman G.H., Carvalho A.F., Matsukuma A., Balza G.S., Simpson D.H.,  
Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare  
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and  
Simpson A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT Laboratory of Cancer Research  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-230  
200-102-d03&cl=2000-02-23&cl=1)  
Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 337.

FEATURES  
Location/Qualifiers  
1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="BN0047"  
/day\_stage="Adult"  
/note="Organ: breast,normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

## alignment\_scores:

Quality: 23.00 Length: 23  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-8 x AM996787/rev ..

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

1 ttleupProaspPrpAaplylspProglNalAlsglntleuglnth 17  
|||||  
105 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAACAGTTGCAGAC 56

17 rARGAlaasPTyrlenu 23  
|||||  
55 CCGTGCAGACTATCTCATC 37

seq\_name: gb\_est1:BB830730

## seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730  
KEYWORDS EST.  
GI:17008973

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 438)  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyu,T.,  
Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL Unpublished (2001)  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyama,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

## FEATURES

## source

Location/Qualifiers  
1..438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930013K04"  
/clone\_11b="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"

## BASE COUNT

138 a 96 c 108 g 96 t

## alignment\_scores:

Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-8 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

1 ttleupProaspPrpAaplylspProglNalAlsglntleuglnth 17  
|||||  
300 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAACAGTTGCAGAC 349

17 rARGAlaasPTyrlenu 22  
|||||  
350 CCGTGCAGACTATCTC 365

seq\_name: gb\_est1:BB834922

## seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930033021 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922  
KEYWORDS EST.  
GI:17013165

SOURCE house mouse.  
ORGANISM Mus musculus

## REFERENCE

1 (bases 1 to 446)  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyu,T.,  
Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL Unpublished (2001)  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Garninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matsubaki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
Riken integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Rikunishi, Y., Shibata, K., Itoh, M., Garninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.  
Location/Qualifiers  
1. 446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G93003J21"  
/clone\_1lb="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) CDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"  
139 a 99 c 108 g 100 t

BASE COUNT  
ORIGIN

Alignment\_scores:  
Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment block:  
US-08-973-363-8 x BB834922 ..

Align seg 1/1 to: BB834922 from: 1 to: 446

1 TleleuPfoApsPpfoApsPlysLysPpfoGlnAlaLysGlnLengthRth 17  
309 ATTCTTCCAGATGATCTTCGTATMAAAACCAAGCCAAACAGTTACAGAC 358  
17 TARGAlaAspTYrleu 22  
|||||  
359 CCCTGCAGACCTACCTC 374

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:  
LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION B01905107F1 NIH\_MGC\_54 Homo sapiens CDNA clone IMAGE:4133129 5',  
mRNA sequence.  
ACCESSION BF239967  
VERSION BF239967.1 GI:11153990  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 821)  
NIH-MGC <http://mgi.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@mai.nih.gov](mailto:cgabs@mai.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLW)

DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LNCMI033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1. 821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1133129"  
 /clone\_id="NH\_MGC\_54"  
 /tissue="bone marrow"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfilI (ggcgccctgcgc); Site\_2: SfilI (ggccatattggc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-ATTCTAGACGCCGCGCCGACATG-dT(30)BN-3'  
 (where B = A, C, G, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t

ORIGIN

alignment\_scores:  
 Quality: 21.00 Length: 21  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x BF239967 ..

Align seg 1/1 to: BF239967 from: 1 to: 821

22 uiletylsleulen 26  
 |||||||  
 53 CATCAAAATTACTT 65

seq\_name: qb\_est1:AW97058

seq\_documentation\_block:  
 LOCUS AW97058 686 bp mRNA linear EST 05-JUN-2000  
 DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW97058  
 VERSION AW97058.1 GI:8257282  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 686)  
 Dias Neto E., Garcia Correa R., Verjovsky-Almeida S., Brites M.R.,  
 Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa P.F.,  
 Goldman G.H., Carvalho A.F., Matsukuma A., Bala G.S., Simpson D.H.,  
 Brunstein A., deOliveira P.S., Bucher P., Jorgensen C.V., O'Hare  
 M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and  
 Simpson A.J.V.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.V.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:  
<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=at2-QV3-BN0047-150>  
 400-152-C03&t3=2000-04-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-8 x AM997058/rev ..  
 Align seg 1/1 to reverse of: AM997058 from: 1 to: 686  
 9 LysPProGAlAlaIysGInleuGInThrArGAlaAspTyrleuGInleuysle 25  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 104 AAACCAAGCAACAGTTCGACGCCGTGCAGACTACTCAATCAAT 55  
 25 uleu 26  
 ||||  
 54 ACTT 51  
 seq\_name: gb\_gss:CNS04DVG  
 seq\_documentation\_block:  
 LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
 103K08 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 AL286261  
 AL286261.1 GI:8024707  
 GSS: genome survey sequence.  
 Tetraodon nigroviridis.  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 856)  
 Roest-Crolius H., Jallou O., Dasilva C., Fizames C., Fisher C.,  
 Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
 Weissenbach J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crolius H., Jallou O., Dasilva C., Bouneau L., Fisher C.,  
 Bernot A., Fizames C., Wincker P., Brothier P., Quetier F.,

Saurin W. and Weissenbach J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/tetraodon>.  
 Location/Qualifiers  
 1..856  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0RG103BF04LP1-end : 77"  
 BASE COUNT 84 a 308 c 252 g 206 t 6 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-8 x CNS04DVG/rev ..  
 Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856  
 2 leuProAspAspProAspLysLysProGInAlaIysGInleuGInThrAr 18  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 674 CTCGCCGATGATCTCTGATAGAAAGCCTCAGGCCAACAGTTACAGCAC 625  
 18 gala 19  
 ||||  
 624 AGCC 621  
 seq\_name: gb\_est2:BF881342  
 seq\_documentation\_block:  
 LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PMO-ET0208-031200-001-f12 ET0208 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS  
 EST.  
 SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Brites M.R.,  
 Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,  
 Goldman G.H., Carvalho A.F., Matsukuma A., Bala G.S., Simpson D.H.,  
 Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare  
 M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and  
 Simpson A.J.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.J.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM0&t=PM0-ET0208-031200-001-f12&t3=2000-12-03&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 128.

FEATURES  
Location/Qualifiers

1..129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung\_tumor; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

us-08-973-363-8 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLenglnThraGAlaAspTyrLeu 22  
|||||  
108 AMGCACTACAGCCGACGCGGATTTACTTG 79

**THIS PAGE BLANK (USPTO)**



COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
Source 1. 1316  
/db\_xref="taxon:32644"  
BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN  
alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x A58696 ..  
Align seg 1/1 to: A58696 from: 1 to: 1316  
1 lleleuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
1138 ATTTCAGTATGATCCAGATTAAGAAACCCAGCGTACAGCAGAC 1187  
17 TATGALASPtyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
1188 TCGTGCAGATTACTCATTAATTACTGAATAAGACCTTGCAAGAAG 1237  
34 lu 34  
||  
1238 AA 1239  
seq\_name: gb\_pat:A58691  
seq\_documentation\_block:  
LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent WO9639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Location/Qualifiers  
Source 1. 6608  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN  
alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608  
1 lleleuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4080 ATTTCAGTATGATCCAGATTAAGAAACCCAGCGTACAGCAGAC 4129  
17 TATGALASPtyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||

4130 CCGTGCAGACTACTCATTAATTACTGAATAAGACCTTGCAAGAAG 4179  
34 lu 34  
||  
4180 AA 4181  
seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University, Glasgow G12 8QQ, UK  
FEATURES Location/Qualifiers  
Source 1. 6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1. 6872  
/gene="CHD-Z"  
228. 5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501846"  
/translation="MNGHSDSESVNRSGSESRSDSDSGASGSGSGSGSGSSGSS  
SOSGSDSESGSESGSSESSTSRERKQVAKPKKADGSEFMKSPSILAVQSAV  
LKRQOQOQKAAASDSDSESDSSSDSADSSSTRTKKKHKDEDMQSGSGSGSGS  
DSESAEDGKSSCESESDYEPKKNKYKSHKPSRLKPKSGKSTGQKROLDSSEE  
DDDEDYDKRGSROATVNVSYKEAEETKTDSDLDLVEGCEVDYPTDEDEFEETIRKMD  
SRIGRKGTAGTSTTYIYAVADGDPPNAGPSEKSELEIOLYLKWKMSLHHNWFEET  
LKOQVKGKANKLDNTKKKDDQETKRWLNKASPDVETVYNCQQLDLDLHKQYIVERTI  
AHSNOKSAGYDYYCKWQGLPYSECSDWDGALIAKKIQARIDETFSRNQSTTPPKD  
CKYLKQRPFRVALRKOPSYTIGHESLELDYQLNOLNMLAHWSGNGSLADENGIG  
KITQITSEFLNYLFHEHOLYGPFLRLSTLSTWSOREIOTMAPONNAVYLIDTSRN  
MIRTHEMHPQTRKLFKNLITLTYELLKDKSFLGLGIMAFVIGVDEARHLKNDLSLY  
RTLIDKSNHRLITGTPLONSIKELMSLHIMPEKSSPEDEEHLKGGVYAS  
LHKLEPELRLRVYKDOEKSLPAKYEQILRMBSLKOQYIKWILTRYKALSKESSG  
STSGFLNIMMELKCCNHCYLIRPPDNEFYKQALDHLIRSSGLILDLKLRLR  
ERGNFVLIFESQVWRMIDILAEYLKROFPFOLDISIGELRKQALDHRNAGSDPFC  
FLISTRAGIGLGINLASADTVVIFSDMNPQNDLQAARAHRIQKQVNIYILVYFGS  
VEEDILERRAKKKVMDHLYVIOEMDTGKTVLHTGTSPPSSPTPEEELAILFGAE  
LEKPEGEQEPQEDIDIELIKRAETRENEGPLVGPDELLSQFVANSNDDEDE  
LEPERNSRWMEETIPESQRRITEERQRELEETIMLRMRCAKOISNGSGRGRSR  
SNRYSGSDSDSITERRKRPKRGRPTIPRENKIGSDAETIRFISYKFKGSGEPRLD  
AARDAEIVDKSETDLRLGELVHNGCIKALDNSGQERAGRGKYGKGPFRISGV  
QVNAKLIVISHEELAPLRLKSISSDEERKRYIYIPOHTAAADIDIMGEDSDNLVGI  
YVYGGSMEMIKMDPLSITOKLIPDDPKRQAOLOTRAYLILKLINKDLAREAO  
RLAGAGNSKRRTKRNKKMKASKIKEELKSSSPQSPSEKSDDEEDDNKEIYSVK  
HLHAKIKTEKNEEKPEPDIGITKAEERETKEKREKREKREKEDDKELKED  
NKEKRENVKSESTOKREVEKERVNEMKSENKESKKIPLDLTPVHTATYSPPVISE  
ESELHOKTFESVCKERMRPVKALKOLDRPEKGLSERQLEHTROCLIKIGDHTECL





[illegible]

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 5947)
JOURNAL	Woodage,T., Basrai,M.A., Baxevanis,A.D., Hieter,P. and Collins,F.S. Characterization of the CHD family of proteins
REFERENCE	Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)
AUTHORS	2 (bases 1 to 5947)
TITLE	Woodage,T.
JOURNAL	Direct Submission
REFERENCE	Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD 20892-4442, USA
FEATURES	Location/Qualifiers
source	1..5947
gene	/organism="Homo sapiens"
CDS	/db_xref="taxon:9606"
	/chromosome="5"
	/map="5q15-21, near WI-5811"
	1..5947
	/gene="CHD1"
	164..5293
	/gene="CHD1"
	/codon_start=1
	/product="CHD1"
	/protein_id="AA87381.1"
	/db_xref="GI:2645429"
	/translation="MNGHSDSEYVRNNGSGSSQSDDDSGASGSGSGSSQSSSDSSSSQSSSDSSGSGSSQSSSESDYFSRKNKYVAKPKYVGAFAEMSSPSILAVQSAIILIKQOQOQOQOQOQOQASNSGSEDSSESSSDSSSEYVRKKKHEDQMGSQSGSPSOSSGSDSEERERKSSCDTESDYEPPKNNVSKPKPNRKSXNGKKILGQKKRQIDSESEDDDEDYDNDKRSRRQATVAVSYKEDDEMKTDSDDLLEVCGEDVPQPEEEFEFTIERFMDCRIGRKGATGATTTIYVAEADGNAGCEKKPEGLIOLIKKGMGSHIHNTWTFEERTKOONRKMKKLDNYKKKQDQETKMKIKASPEDEYVYNCQOELIDDLIKOYQIYGRITAHMSKSAQIPDYCKMQGLPYSECSEWDGALSKFQACIDETFSRNSKTTPRFKDCKLQKRRPRFVALRKQPSYIGGHEGLERDQLGNLAWSCNGNSCILLADEMGLGCTIOTISFLNLFHEHOLYGPFLVLPSTLSMQRLOTWASOMNAVYLGIDINL SRNRIPTHETWHQTKRLKFNILITTYEILDKAPLGLMWARIGYDEARLKNDSYSLAYTLIDFSKNHLITGTPLONSLKEMLGLHFIPEKFSWDEFEHGGKREYGA VASIKELPEPILBRVKRDKVSLPAKVEOILRMESMLQOYKWLITRNYKALSGY SKGTSQGLNIMELKKCNHCYILIRPPDNNEFYNGKALDILRSSGKITLIDLLILIRLRGNRVILFSQYMYMDILATYLYLKYRPPORIDGSIKGEIRKQALDHFNAEGSERDFCEFLSTRAGGIGLINASADTVIFEDSMQPDLDQAQARHRTGQKQOVNTRIVYKGSVEEDILERAKRKWLDLVIQRMOTYTKTVLHTSAPSSSTPPKKEEYALIKRFGAEELFKPEGEDEPQEMQDIDELIKRAETHENEGPLVGBELLQKRVANFSNMDEDDIELEPBRNSKNMEIIPEDORRLKEEEROKELEIETMLPRMCAKQLSFNGSESRSRNRYSGSDSDSISEGKRPRKRGRTIPRENIKGFSDAIEIRPLTISKYKFGGPELRLDILARDAELVDSSTPDRLRGLVHNGCTIKALKDSSGTERGGLGYKKCFPHITSGOVYNAKLVISHEEELILPKRSTIPSPBERKQTTICHTKRAAFIDIMGKEDDSNLLIGIYEVYGSWEMTKMPDLSLTHRLILPDDPKRPOAKQIOTRADYILIKLSRLARK EALSGASSSKRRKARAKNNKAMKSIKVEELKSQSSPLSPKRSQEDDKLSEKSDSR ERSKSSVSDAPVHITASGEVPVISESEELQDFTSICKERMVPAAALQOLDPPEPKR GLSREDEHTRQCLIKIGDITCELEKYTNPEOIKQMRKLMIFKFTFEFDARKLHKLYHAKIKRQESQONDNSNLNPHYIRNDVPERLKNTHHDSSDSYSDSHRLKRYHDKRHQCDYKKSQDSKRRPYSNGCKDHDQMHYQDQSYFYDREHRRKLDLP RSRDRHNRNLSGLKDRSHSDHSHSDHLSHDHSSSYTHHSSRQYRTHSDQMOMDHR RASSSGRSPIDQSSYSGSRSPFSEHSHKSTPEHTWSSRKT"

```
1 11leuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17
|||||
4025 ATTCTTCCAGATGATCCGATTAACCAACGACAAAGAGTTGCAGAC 4074
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
4075 CCGTGCAGACTACCTCATCAATTAATTCTT 4102
seq_name: gb_pr:AC092372

seq_documentation_block:
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
ACCESSION AC092372
VERSION AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:1529048.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 146.7kb). It is clipped at the overlap with AC012624.
The number of bases overlapped is 90404.
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-58M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AC092372/rev ..
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220
1 11leuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17
|||||
26932 ATTCTTCCAGATGATCCGATTAACCAACGACAAAGAGTTGCAGAC 26883
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
26882 CCGTGCAGACTACCTCATCAATTAATTCTT 26855
seq_name: gb_pr:AC012624
```

```
seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
1..134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365
1 11leuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17
|||||
118247 ATTCTTCCAGATGATCCGATTAACCAACGACAAAGAGTTGCAGAC 118296
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
118297 CCGTGCAGACTACCTCATCAATTAATTCTT 118324
seq_name: gb_htg:AC021449

seq_documentation_block:
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 143079)
```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-58M12  
Unpublished  
2 (bases 1 to 143079)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,  
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,  
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
DeRellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Lander, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGuirk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Meus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced g1:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L5154  
Center clone name: 58\_M12

Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q20  
Consensus quality: 140814 bases at least Q20  
Insert size: 144000; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 10 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length

\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 38820  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"

misc\_feature  
38921..40411  
/note="assembly-fragment"  
40512..43279  
/note="assembly-fragment"

misc\_feature  
43380..46905  
/note="assembly-fragment"  
47006..51830  
/note="assembly-fragment"

misc\_feature  
51931..62619  
/note="assembly-fragment"  
62720..75408  
/note="assembly-fragment"

misc\_feature  
75509..92516  
/note="assembly-fragment"  
92617..106409  
/note="assembly-fragment"

misc\_feature  
106510..143079  
/note="assembly-fragment  
clone\_end:17  
vector\_side:right"

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC021449 ..

Align seg 1/1 to: AC021449 from: 1 to: 143079

1 HleuProAspAspProAspLysProGlnAlaLysGlnLeuInth 17  
|||||  
116147 ATTCTTCCAGATGATCCCGATAAACCAACGCAACGATTGCAGAC 116196

17 FARGAlAspTyrLeuIleLysLeuLeu 26  
|||||  
116197 CCGTGCAGACTACTCTCAATTAATTACTT 116224

seq\_name: gb\_hrg:AC008531

seq\_documentation\_block:  
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.

ACCESSION AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE  
2 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 14, 2001 this sequence version replaced g1:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 369535  
Center clone name: CRT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000: sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
-----  
1 56174: contig of 56174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 118180: contig of 4963 bp in length  
\* 118181 118280: gap of unknown length  
\* 118281 118290: contig of 1404 bp in length  
\* 118291 119694: gap of unknown length  
\* 119695 123297: contig of 3503 bp in length  
\* 123298 123397: gap of unknown length  
\* 123398 145659: contig of 22262 bp in length.  
-----  
Location/Qualifiers  
1 145659  
/organism="Homo sapiens"  
/dbxref="taxon:9606"  
/chromosome="5"  
/clone\_lib="catrech human BAC library C"  
/clone\_id="26309 c 27380 g 48609 c 600 others  
-----  
BASE COUNT 42561 a 26309 c 27380 g 48609 c 600 others  
ORIGIN  
-----  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
-----  
alignment\_block:  
US-08-973-363-9 x AC008531 ..  
Align seg 1/1 to: AC008531 from: 1 to: 145659  
-----  
1 HleuPProAspAspProAspLyseProGlnAlaIysGlnLeuGlnH 17  
|||||  
46770 ATTCTCCAGATGATCCGATATAAAACCAACGAAACAGTTGCAGAC 46819  
|||||  
17 TATGTAASPTVTLGIIETLSLEULEU 26  
|||||  
46820 CCGTCGAGACTACCTCATCAATAATTAATT 46847  
|||||  
seq\_name: gb\_hlg:AC091946

seq\_documentation\_block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (03-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCT-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590: agarose-fp estimation  
Estimated insert size: 190246: sum-of-contigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----  
1 1358: contig of 1358 bp in length  
\* 1359 1439: gap of unknown length  
\* 1439 2667: contig of 1209 bp in length  
\* 2667 2767: gap of unknown length  
\* 2768 2823: contig of 1056 bp in length  
\* 2823 3824: gap of unknown length  
\* 3824 3923: gap of unknown length  
\* 3923 5920: contig of 1997 bp in length  
\* 5920 6020: gap of unknown length  
\* 6020 7494: contig of 1474 bp in length  
\* 7494 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12450: contig of 2100 bp in length  
\* 12451 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14711 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18068: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20534: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 27563: gap of unknown length  
\* 27563: contig of 3948 bp in length

```

* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57460: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74133: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.

```

FEATURES  
source 1..193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_id="RP11-36012" RPI1 human BAC library 11"

BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC091946 ..

Align seg 1/1 to: AC091946 from: 1 to: 193446

```

1 ttleuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnth 17
|||||
87658 ATTCTTCGAGATGCCGATATAAAACCAAGCAAGCAAGTTCGAGAC 87707
17 rArgAlaAspTyrLeuLeuLysLeuLeu 26
|||||
87708 CCGTGCAGACTACCTCATCAATTAATTACTT 87735

```

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:  
LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.  
ACCESSION AC026778

```

VERSION AC026778.4 GI:14277282
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 195433)
JOURNAL Direct Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Unpublished
TITLE 2 (bases 1 to 195433)
JOURNAL DOE Joint Genome Institute.
REFERENCE Direct Submission
AUTHORS Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL 3 (bases 1 to 195433)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
TITLE Drive, Walnut Creek, CA 94598, USA
JOURNAL On Jun 1, 2001 this sequence version replaced gi:13677045.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
STS Content:
WT-13675 G23101
SHGC-58345 G38487
SHGC-103595 G57841.

```

FEATURES  
source 1..195433  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC026778/rev ..

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

```

1 ttleuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnth 17
|||||
29732 ATTCTTCGAGATGCCGATATAAAACCAAGCAAGCAAGTTCGAGAC 29683
17 rArgAlaAspTyrLeuLeuLysLeuLeu 26
|||||
29682 CCGTGCAGACTACCTCATCAATTAATTACTT 29655

```

seq\_name: gb\_pr:AC022121

seq\_documentation\_block:

LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001  
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
ACCESSION AC022121  
VERSION AC022121.6 GI:15375145

KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 219258)  
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Title  
JOURNAL Direct Submission (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
REFERENCE Direct Submission  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Title  
JOURNAL Direct Submission  
Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
JOURNAL Title  
JOURNAL Direct Submission  
Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced g1:15148108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
SPS Content:  
WT-5811 GM4974  
WT-13675 G23101  
SHGC-58345 G38467  
SHGC-103595 G37841.

FEATURES  
source Location/Qualifiers  
1..219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"

BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC022121/rev ..

Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258

1 lleleuproaspaprophaslystysproglalalyglnleuglnrh 17  
|||||  
89450 ATTCTTCAGATGATCCGATATAAAACACACAGCAACAGATTGCAGAC 89401  
17 TAAGAlasPTytleuUleUleUleUleUleUleUleUleUleU 26  
|||||  
89400 CGGTGAGAGTACTCTCATCAATTAATTCTT 89373







```

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24
|||||
52 AAGAAACCAAGCAAGCAAGTTCAGACCCGTCAGATTACTCATTTAA 101
|||||
24 sleuLeuAsnLysAspleuAlaArgLysGlnValGlnArgLeuThrGlyA 41
|||||
102 ATTACTGTAATTAAGACCTTCAGAAAGAAAGACTCAAGACTTACTGTT 151
|||||
41 la 41
||
152 CA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142754
seq_documentation_block:
ID AA142754 standard; cDNA; 1316 BP.
XX
AC AA142754;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chicken CHD-W gene (partial sequence).
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
OS Gallus sp.
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI; 1997-043127/04.
XX
PT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
PS Claim 1; Fig 8; 76pp; English.
XX
CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with
CC the closely related CHD-1A gene (AA142751) is suggested to initiate
CC female development in birds. The sequence of CDH-W was deduced
CC from 2 clones isolated from a 10-day chick embryo library using
CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is
CC located on the W chromosome. Probes based on CHD-W and CHD-1A give
CC a W chromosome-specific signal on hybridisation to genomic DNA of a
CC non-rare bird and can be used for sex determin. of a bird. CHD-W
CC nucleic acids can also be used to control the sex of progeny of a
CC bird.
XX
SO Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AA142754 ..
Align seg 1/1 to: AA142754 from: 1 to: 1316
1 ||leuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThr 17
|||||

```

```

1138 ATTTTACCTGATGATCCAGATAGAAACCCAGGCTAAGCATTACAGAC 1187
|||||
17 FA14AlaAspTyrLeuIleLysleuAsnLysAspleuAlaArgLysG 34
|||||
1188 TCGTGCAGATTACTCATTTAATTACTGTAATTAAGACCTTCAGAAAGG 1237
|||||
34 lu 34
||
1238 AA 1239

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751
seq_documentation_block:
ID AA142751 standard; cDNA; 6608 BP.
XX
AC AA142751;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chicken CHD-1A gene.
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT CDS 228..5390
FT /*tag= a
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI; 1997-043127/04.
XX
DT Avian chromodomain-helicase-DNA binding genes determine sex in
DT birds - used for sex determ. and to control sex of progeny
XX
PS Claim 1; Fig 5; 76pp; English.
XX
CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with
CC the closely related CHD-1A gene (AA142751) is suggested to initiate
CC female development in birds. The sequence of CDH-1A was deduced
CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA
CC library using a great tit CHD-W sequence (see also AA142755) as probe.
CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1
CC gene (see also AA142756-57). It is located on an autosome or Z
CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-
CC specific signal on hybridisation to genomic DNA of a non-rare
CC bird and can be used for sex determin. of a bird. CHD-1A nucleic
CC acids can also be used to control the sex of the progeny of a bird.
XX
SO Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AA142751 ..
Align seg 1/1 to: AA142751 from: 1 to: 6608

```

```

1 11leuProaspaspProaspLySpProGlnAlaLySGlnLeuGlnH 17
|||||
4080 ATTTACTGTCATGATCCAGACAAAGCCAGCAAGCAGTACAGAC 4129
|||||
17 rrrglaaspyrleuilelyleuanslysaspleuAlaArglysg 34
|||||
4130 CCGTCACACACCTCATTAATTAATTAATTAAGACCTTCAGAAAGG 4179
|||||
34 lu 34
|||||
4180 AA 4181

```

seq.name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AA142757

```

seq.documentat_block:
ID   AA142757 standard; DNA; 153 BP.
AC   AAT42757;
XX
XX
DT   12-MAR-1997 (first entry)
XX
DE   Chick CHD-1A gene fragment.
XX
XX
KW   Bird; sex determination; chromodomain-helicase-DNA binding 1;
KM   CHD-1A; CHD-W; W chromosome; ss.
XX
OS   Gallus sp.
XX
FH   Key
FT   misc_difference 52..81
FT                                     /tag= a
FT                                     /note= "bases 52-81 are a repeat of bases 22-51
FT                                     and are ignored in the translated amino
FT                                     acid sequence given in Fig 3"
XX
XX
PN   WO9639505-A1.
XX
PD   12-DEC-1996.
XX
PF   05-JUN-1996; 96WO-GB01341.
XX
PR   06-JUN-1995; 95GB-0011439.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Griffiths R, Tiwari B;
XX
PI   Griffilhs R, Tiwari B;
XX
DR   WPI: 1997-043127/04.
DR   P-PSDB: AAM08147.
XX
XX
PT   Avian chromodomain-helicase-DNA binding genes determine sex in
PT   birds - used for sex determ. and to control sex of progeny
XX
XX
PS   Claim 8; Fig 3; 76pp; English.
XX
XX
CC   Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC   to portions of the chicken CHD-1A (A - Avian) gene (AA142757),
CC   chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)
CC   and the great tit CHD-W gene (see also AA142759). Translated amino
CC   acid sequences of this region are provided in AAM08146-49. The
CC   CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes
CC   determine sex in birds and can be used to identify the sex of an
CC   embryo, foetus etc. and to manipulate the sex of progeny.
XX
XX
SQ   Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

```

alignment\_scores:      Quality: 27.00      Length: 27  
                              Ratio: 1.000      Gaps: 0  
 Percent Similarity: 100.000      Percent Identity: 100.000

alignment\_block:      us-08-973-363-9 x AAT42757

Align seg 1/1 to: AAT42757 from: 1 to: 153

```

8 lvslyProGlnAlaLySGlnLeuGlnHrrrglaaspyrleuilely 24
|||||
52 AAGAAACCCAGCAAGCAAGCAGTACAGACCGTCAGACTCATTA 101
|||||
24 slleuanslysaspleuAlaArglysglu 34
|||||
102 ATTTACTGATTAAGACCTTCACAGAAAGGAA 132

```

seq.name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AA142758

```

seq.documentat_block:
ID   AAT42758 standard; DNA; 153 BP.
XX
XX
AC   AAT42758;
XX
XX
DT   12-MAR-1997 (first entry)
XX
DE   Chick CHD-W gene fragment.
XX
XX
DE   Chick CHD-W gene fragment.
XX
XX
KW   Bird; sex determination; chromodomain-helicase-DNA binding 1;
KM   CHD-1A; CHD-W; W chromosome; ss.
XX
XX
OS   Gallus sp.
XX
FH   Key
FT   misc_difference 52..81
FT                                     /tag= a
FT                                     /note= "bases 52-81 are a repeat of bases 22-51
FT                                     and are ignored in the translated amino
FT                                     acid sequence given in Fig 3"
XX
XX
PN   WO9639505-A1.
XX
PD   12-DEC-1996.
XX
PF   05-JUN-1996; 96WO-GB01341.
XX
PR   06-JUN-1995; 95GB-0011439.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Griffiths R, Tiwari B;
XX
PI   Griffilhs R, Tiwari B;
XX
DR   WPI: 1997-043127/04.
DR   P-PSDB: AAM08148.
XX
XX
PT   Avian chromodomain-helicase-DNA binding genes determine sex in
PT   birds - used for sex determ. and to control sex of progeny
XX
XX
PS   Claim 8; Fig 3; 76pp; English.
XX
XX
CC   Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC   to portions of the chicken CHD-1A (A - Avian) gene (AA142757),
CC   chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
CC   and the great tit CHD-W gene (AA142759). Translated amino acid
CC   sequences of this region are provided in AAM08146-49. The CHD-1A
CC   (see also AA142751) and CHD-W (see also AA142754-55) genes determine
CC   sex in birds and can be used to identify the sex of an embryo,
CC   foetus etc. and to manipulate the sex of progeny.
XX
XX
SQ   Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

```

alignment\_scores:      Quality: 27.00      Length: 27  
                              Ratio: 1.000      Gaps: 0  
 Percent Similarity: 100.000      Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x AAT42758 ..

Align seg 1/1 to: AAT42758 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrIleuIlely 24

|||||

52 AAGAAACCCAGCGCTTACAGACCCGTCAGATTACCTCATTTAA 101

24 sLeuLeuAsnLysAspLeuAlaArgLysGlu 34

102 ATTACTGTAATTAAGACCTTCAGCAAGAAAGGA 132

seq\_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280

## seq\_documentation\_block:

ID AAV59280 standard: cDNA: 1311 BP.

XX AAV59280;

XX 14-DEC-1998 (first entry)

XX DE Altered telomere repeat binding factor 1 gene.

XX ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;

XX telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

XX Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 1..1311

XX FT /\*tag= a

XX FT /product= "A-TRF"

XX PN WO9836066-A1.

XX PD 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US02765.

XX PR 04-FEB-1998; 98US-0018628.

XX PR 13-FEB-1997; 97US-0800264.

XX PA (UYRQ ) UNIV ROCKEFELLER.

XX PI Bianchi A, De Lange T, Van Steensel B;

XX DR WPI: 1998-480769/41.

XX DR P-PSDB: AAV59280.

XX PT Nucleic acid encoding altered telomere repeat binding protein and

XX PT related vectors - transformants, hetero-dimers and antibodies, used

XX PT to inhibit shortening of telomerases caused by ageing or disease,

XX PT also used to extend life of cells in culture

XX PS Claim 14; Page 110-111; 163pp; English.

XX CC The altered vertebrate telomere repeat binding protein (A-TRF) has a

XX CC telomere repeat binding factor (TRF) dimerisation domain, and forms a

XX CC hetero-dimer with TRF, preventing it from binding to the specified repeat

XX CC sequence. A-TRF, optionally expressed by gene therapy, is used to

XX CC inhibit shortening of telomeres associated with ageing (for cosmetic

XX CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,

XX CC atrophy of the skin, age-related macular degeneration, atherosclerosis,

XX CC tumours and viral (including human immune deficiency virus) infection.

XX CC Cells expressing A-TRF also have an increased life span in vitro, e.g.

XX CC for expression of recombinant proteins or where intended for subsequent

XX CC transplant or for testing, eliminating the need for transformation.

XX CC Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

## alignment\_scores:

Quality: 25.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

US-08-973-363-9 x AAV59280 ..

Align seg 1/1 to: AAV59280 from: 1 to: 1311

1 lLeuLeuProAspPheProAspLysLysProGlnAlaLysGlnLeuGlnTh 17

1237 ATTCTCCAGATGATCCTGATTAACCAACCAACCAACGTTACAGAC 1286

17 rArgAlaAspTyrIleuIleLysIleu 25

1287 CCGTGCAGACTACTCTCATCAACTTA 1311

seq\_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAV42756

## seq\_documentation\_block:

ID AAT42756 standard: DNA: 153 BP.

XX AAT42756;

XX 12-MAR-1997 (first entry)

XX DE Mouse CHD-1 gene (bases 3855-9777).

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1; CHD-W; W chromosome; ss.

XX Mus sp.

XX Key Location/Qualifiers

XX FT misc\_difference 52..81

XX FT /\*tag= a

XX FT /note= "bases 52-81 are a repeat of bases 22-51

XX FT and are ignored in the translated amino

XX FT acid sequence given in Fig 3"

XX PN WO9639505-A1.

XX PD 12-DEC-1996.

XX PF 05-JUN-1996; 96WO-GB01341.

XX PR 06-JUN-1995; 95GB-0011439.

XX PR (ISIS-) ISIS INNOVATION LTD.

XX PA Griffiths R, Tiwari B;

XX PI WPI: 1997-043127/04.

XX DR P-PSDB: AAW08146.

XX PT Avian chromodomain-helicase-DNA binding genes determine sex in

XX PT birds - used for sex determin. and to control sex of progeny

XX PS Claim 8; Fig 3; 76pp; English.

XX CC Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology

XX CC to portions of the chicken CHD-1A (A = Avian) gene (see also

XX CC and AAT42757). chicken CHD-W (W refers to the W chromosome) gene

XX CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).

XX CC Translated amino acid sequences of this region are provided in

XX CC AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also

XX CC AAT42754-55) genes determine sex in birds and can be used to identify

XX CC the sex of an embryo, fetus etc. and to manipulate the sex of

XX CC progeny.

XX CC Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

alignment\_scores:                   Quality: 19.00                   Length: 19  
                                      Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AAT42756       ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

8 LysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAAAAACACACAGCAAAACAGTACAGACCCGTCGACACTACCTCATCA 101  
|||||  
24 sleuieu 26  
102 ACTACTT 108

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:  
ID ABL06443 standard; cDNA: 6240 BP.  
XX  
AC ABL06443:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;  
DR P-PsDB; ABB62340.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

alignment\_scores:

alignment\_scores:                   Quality: 11.00                   Length: 11  
                                      Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x ABL06443       ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
4205 AAGCCACAGCCAGCAGCGTCGACGCGTCC 4237  
|||||

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:  
ID ABL06442 standard; cDNA: 9933 BP.  
XX  
AC ABL06442:  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;  
DR P-PsDB; ABB62339.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

alignment\_scores:

Quality: 11.00                   Length: 11  
Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x ABL06442       ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933  
9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
6745 AAGCCCAAGGCCAAGCAGCTGCAGACCGTGC 6777  
seq\_name: /SIDSI/gcgdata/hold-geneseq/genesegn-emb1/NA2001A.DAT:AAK88882  
seq\_documentation\_block:  
ID AAK88882 standard: cDNA: 421 BP.  
XX  
AC AAK88882;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226861.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232081.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-502630/55.  
XX P-PSDB: AAM93109.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
XX PT diagnosing, treating, preventing and/or prognosing disorders of the  
XX PT digestive system, particularly cancer and cancer metastases -  
XX  
XX PS Claim 1; SEQ ID NO 1199; 986pp; English.  
XX  
XX CC The present invention provides the protein and coding sequences of a  
XX CC number of human digestive system antigens. These can be used in the  
XX CC diagnosis, treatment and prevention of digestive system disorders,  
XX CC including cancer, Meckel's diverticulum, bacterial or parasitic  
XX CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
XX CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
XX CC system antigen of the invention.  
XX  
XX SO Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-9 x AAK888882 ..  
  
Align seg 1/1 to: AAK888882 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrIleu 22  
|||||  
146 AACGAGCTACGACCGACCGGATTAATTG 175  
  
seq\_name: /SIDS1/gc9data/hold\_geneseg/genesegn-emb1/NA2001A.DAT:AAI57603  
seq\_documentation\_block:  
ID AAI57603 standard; cDNA: 421 BP.  
XX  
XX AAI57603;  
XX  
XX  
XX 19-OCT-2001 (first entry)  
XX  
XX DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.  
XX  
XX KM Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
XX  
XX OS Homo sapiens.

XX  
XX WO20015350-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01350.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189876.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226581.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227099.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231245.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0232403.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
XX 25-SEP-2000; 2000US-0234997.  
XX 25-SEP-2000; 2000US-0234998.  
XX 26-SEP-2000; 2000US-0235484.  
XX 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-457727/49.  
DR P-PSDB; AAM38625.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon and rectum including colorectal cancers  
PR and also for testing and detection e.g. diagnosis -  
XX  
PS Claim 1: SEQ ID NO: 67; 522pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in  
CC AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the  
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
CC The present sequence is a colorectal cancer antigen coding sequence of  
CC the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:                   Length: 10  
                  Quality: 10.00                   Gaps: 0  
                  Ratio: 1.000                   Percent Identity: 100.000  
Percent Similarity: 100.000                   Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-9 x AAI57603 ..  
  
Align seg 1/1 to: AAI57603 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
PR 146 AAGCAGCTACAGACCGAGCGGATTACTTG 175  
  
seq\_name: /stidsl/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV33521  
seq\_documentation\_block:  
ID AAV33521 standard; cDNA; 252 BP.  
XX  
XX AAV33521;  
AC  
XX  
DT 29-DEC-1998 (first entry)  
XX  
DE Clone 23789 cDNA fragment encoding a human protein kinase homolog.  
XX  
XX Protein kinase; cell signalling; inflammation; carcinoma; diabetes;  
KW human X-linked agammaglobulinemia; nonspherocytic haemolytic anaemia;  
KW artherosclerosis; glioma; restenosis; cholera-based septic shock;  
KW CHKFRNK chicken tyr kinase; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX US5817479-A.  
PN  
XX  
XX 06-OCT-1998.  
PD  
XX  
XX 07-AUG-1996; 96US-0700575.  
PF  
XX  
XX 07-AUG-1996; 96US-0700575.  
PR  
XX  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX  
XX Au-Young J, Bandman O, Hawkins PR, Wilde CG;  
XX  
XX WPI: 1998-556387/47.  
DR  
XX  
XX Human kinase polynucleotide(s) and recombinant products - useful  
PT for identification of modulators of the enzyme, and treatment of  
PT diseases associated with abnormal kinase expression  
PT





```
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-973-363-9 x AAK10214  ..

Align seg 1/1 to: AAK10214 from: 1 to: 544

21 TyrleuilelyslleuAsnlys 28
  ||||||||||||||||||||
243 TACCTGATTAAGTATTAAATAAG 266

seq_name: /SIDS1/gcgdata/hol1-geneseq/geneseqn-emb1/NA2001A.DAT:AAK36112

seq_documentation_block:
ID AAK36112 standard; DNA; 544 BP.
XX
AC AAK36112;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.
XX
KM Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633666.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 10669; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
```

```
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-973-363-9 x AAK36112  ..

Align seg 1/1 to: AAK36112 from: 1 to: 544

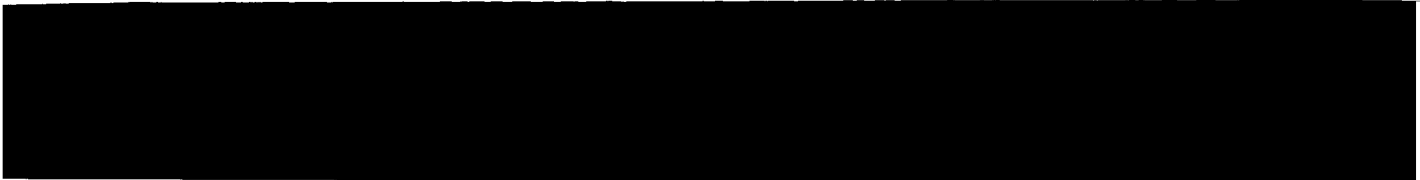
21 TyrleuilelyslleuAsnlys 28
  ||||||||||||||||||||
243 TACCTGATTAAGTATTAAATAAG 266
```

Mon Aug 5 11:52:02 2002

us-08-973-363-9.olip2n.rng

---

Page 11



**THIS PAGE BLANK (USPTO)**



DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone l1e1d12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoes; Pipidae; Xenopodine; Silurana.  
REFERENCE 1 (bases 1 to 645)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
CONTACT: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropesanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: l1e1d12.pic  
Sequencing primer: p1c  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
FEATURES  
source  
1..645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="l1e1d12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"  
BASE COUNT 222 a 125 c 156 g 141 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 30.00 Length: 30  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x AL644594 ..  
Align seg 1/1 to: AL644594 from: 1 to: 645  
1 l1e1euProaspPProaspPlyslsProGlnAlaLysGlnLeuGlnTh 17  
|||||  
456 ATTTCACCATGATCCAGATAGAACCCCAAGCCAGCAGCAGAC 505  
17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeu 30  
|||||  
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGAATCTG 545  
seq\_name: gb\_est1:AL601246  
seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFZP313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKFZP313J1040 5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
Bloeker,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and Wiemann  
,S.

TITLE EST (Bloeker,H., Boecher,M., Brandt,P., Mewes,H.W., Well,B. and Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No sl sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.  
FEATURES  
source  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB; cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x AL601246 ..  
Align seg 1/1 to: AL601246 from: 1 to: 430  
1 l1e1euProaspPProaspPlyslsProGlnAlaLysGlnLeuGlnTh 17  
|||||  
218 ATTCTTCAGATGATCCGATTAATAAACCAACCAAGCAGTTCAGAC 267  
17 rArgAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
268 CCGTGCAGACTACCTCATCAATTAATTCTT 295  
seq\_name: gb\_est1:A1890775  
seq\_documentation\_block:  
LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wm95f11.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; , mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.



1 l1leuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
489 ATCTTCACAGATGATCCTGATTAACCAACCAACAAAGATTACAGAC 538  
17 lFARAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
539 CCGTGCAGACTACCTCATCAACTACTT 566  
seq\_name: gb\_est1:BB461065  
seq\_documentation\_block:  
LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to l10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
ACCESSION BB461065  
VERSION BB461065.2 GI:16426612  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 660)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeba,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (2001)  
Unpublished (2001)  
TITLE On Jul 21, 2000 this sequence version replaced gi:9356558.  
JOURNAL Contact: Yoshhide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
FEATURES  
Source location/Qualifiers  
1. .660

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Saito, Site 2: BamHI, cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGCGCCGCACTCGAGTTTCTTTTCTTTTCTTTVN 3']. cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGATTCGAGTTTAAATTAATTAATCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x BB461065 ..  
Align seg 1/1 to: BB461065 from: 1 to: 660

1 l1leuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
481 ATCTTCACAGATGATCCTGATTAACCAACCAACAAAGATTACAGAC 530  
17 lFARAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
531 CCGTGCAGACTACCTCATCAACTACTT 558  
seq\_name: gb\_est1:AU125712  
seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isogai,T.)  
Unpublished (2000)  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp



HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

## FEATURES

Location/Qualifiers

1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_1lb="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AU125712

Align seg 1/1 to: AU125712 from: 1 to: 866

1 11leuProaspAspProaspLysLysProGlnAlaLysGlnLeuGlnh 17  
|||||  
450 ATCTCTCCAGATGATCCGATTAACCAAGCAAAAGATTGCGAC 499  
|||||  
17 TARGAlaAspTyrLeuLeuLysLeuLeu 26  
|||||  
500 CCGTCAGACATCACTCAATTAATCTT 527

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060P1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.

ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1028)  
NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM9753 row: h column: 16

High quality sequence stop: 488.

Location/Qualifiers

1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_1lb="NIH\_MGC\_72"  
/tissue\_type="melanocytic melanoma"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: skin; Vector: PCMV-SPORT6; Site:1; Notti;

Site:2: Sall; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x BE895133

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 11leuProaspAspProaspLysLysProGlnAlaLysGlnLeuGlnh 17  
|||||  
131 ATCTCTCCAGATGATCCGATTAACCAAGCAAAAGATTGCGAC 180  
|||||  
17 TARGAlaAspTyrLeuLeuLysLeuLeu 26  
|||||  
181 CCGTCAGACATCACTCAATTAATCTT 208

seq\_name: gb\_est1:AW96787

seq\_documentation\_block:

LOCUS AW96787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-003 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW96787  
VERSION AW96787.1 GI:8257021  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 337)

NIH-MGC

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM9753 row: h column: 16

High quality sequence stop: 337.

Location/Qualifiers

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="BN0047"  
/dev\_stage="Adult"  
/note="Organ: Breast; normal; Vector: puc18; Site:1; SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

Profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

#### alignment\_scores:

Quality: 23.00 Length: 23  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-9 x AW996787/rev ..

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

1 ltleuProaspPpRoaspLysPpRoGlnAlaLysGlnLeuGlnth 17  
|||||  
105 ATTCTTCAGATGATCCGATATAAACACAAAGCAAGATTGCAGAC 56

17 rArgAlaAspTyrLeu 23  
|||||  
55 CCGTGACAGACTATCTCTC 37

seq\_name: gb\_est1:BB830730

#### seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730

VERSION BB830730.1 GI:17008973

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 438)

AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,U., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

URL: genome-resgsc.riken.go.jp

Normalisation and subtraction of cap-trapped-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

, Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

e mouse tissues.

FEATURES  
Location/Qualifiers

source

1..438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

/clone\_id="G930013K04"  
/clone\_11b="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"

/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT 138 a 96 c 108 g 96 t

ORIGIN

#### alignment\_scores:

Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-9 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

1 ltleuProaspPpRoaspLysPpRoGlnAlaLysGlnLeuGlnth 17  
|||||  
300 ATTCTTCAGATGATCCGATATAAACACAAAGCAAGATTGCAGAC 349

17 rArgAlaAspTyrLeu 22  
|||||  
350 CCGTGACAGACTATCTCTC 365

seq\_name: gb\_est1:BB834922

#### seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.

ACCESSION BB834922

VERSION BB834922.1 GI:17013165

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 446)

AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,U., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

URL: genome-resgsc.riken.go.jp



Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150>)  
 400-152-c036t3-2000-04-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers  
 1. 686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"

/note="Organ: Breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t

ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-9 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

9 LysPProGlnAlaLysGlnLeuGlnThrArgAlaLysPtyrLeuIleLysLe 25  
 |||||  
 104 AAGCCACAGCAAAACAGTTCCAGACCCGTCGACACTACCTCACTCAATT 55  
 25 uleu 26  
 |||||  
 54 ACTT 51

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:  
 LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
 103608 of library G from Tetradon nigroviridis, genomic survey  
 sequence.  
 AL286261  
 VERSION AL286261.1 GI:8024707  
 GSS: genome survey sequence.  
 KEYWORDS Tetradon nigroviridis.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Acanthomorphae; Acanthopterygii; Perciformae; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigroviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.  
 Location/Qualifiers  
 1. 856  
 /organism="Tetradon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BG103BF04LPI-end : T7"

BASE COUNT 84 a 308 c 252 g 206 t 6 others

ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-9 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 LeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThr 18  
 |||||  
 674 CTGCCGATGATCTCTGATAGAGACCTCAGGCCAAGCATTACAGCAC 625  
 18 gala 19  
 |||||  
 624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:  
 LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PM0-ET0208-031200-001-f12 ET0208 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 CONTACT: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project: This entry can be seen in the following URL:  
<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM0&t2=PM0-ET0208-031200-001-t12&t3=2000-12-03&t4=1>

Seq primer: puc 18 forward  
 High quality sequence start: 22  
 High quality sequence stop: 128.

# FEATURES

## Source

1. 129  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="ET0208"  
 /dev\_stage="Adult"  
 /note="Organ: lung\_tumor; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 21 a 48 c 23 g 37 t  
 ORIGIN

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-9 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLeuGlnThrArgAlaSPtyLeu 22  
 ||||||||||||||||||||||||||||  
 108 AAGCAGCTACAGACCGACGCGATTACTTg 79

**THIS PAGE BLANK (USPTO)**



Query Match 100.0%; Score 6608; DB 6; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCTGGGCGACGAAAGCGACCCGCGGCGACGCGAGGCTGGGGGGAAGGCTGGC 60  
Db 1 CGGGCTGGGCGACGAAAGCGACCCGCGGCGACGCGAGGCTGGGGGGAAGGCTGGC 60

QY 61 CGGGCGAGCCGCGAGCGACGAGGTATTTGGCCAAAATCTTGCCATCTGTAGAGAAATG 120  
Db 61 CGGGCGAGCCGCGAGCGACGAGGTATTTGGCCAAAATCTTGCCATCTGTAGAGAAATG 120

QY 121 CAACTCAAGCATTACTTCGAAAACATAGCGAGTACGAGAAAGGGGATCTTGACCTAC 180  
Db 121 CAACTCAAGCATTACTTCGAAAACATAGCGAGTACGAGAAAGGGGATCTTGACCTAC 180

QY 181 ACCTTGTAACCTGAGTGGAGCTTTCTTTTAACCTTTAACTTAACTTACATGATGGGACA 240  
Db 181 ACCTTGTAACCTGAGTGGAGCTTTCTTTTAACCTTTAACTTAACTTACATGATGGGACA 240

QY 241 GTGATGAAGAAGTGTAAAGAAACAGACGTGAGAGTCAAGCATCAGATGATGATCTGTG 300  
Db 241 GTGATGAAGAAGTGTAAAGAAACAGACGTGAGAGTCAAGCATCAGATGATGATCTGTG 300

QY 301 GGTGAGCTTCAGGTTCTGATCTGGTTCAGCTCTGGAAGCAGTACGAGTGAAGTACCA 360  
Db 301 GGTGAGCTTCAGGTTCTGATCTGGTTCAGCTCTGGAAGCAGTACGAGTGAAGTACCA 360

QY 361 GCCAGTCAGTGAAGTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTG 420  
Db 361 GCCAGTCAGTGAAGTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTG 420

QY 421 AGTGTGACACATCTAGAGAGAGAAAGAAAGTCAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 480  
Db 421 AGTGTGACACATCTAGAGAGAGAAAGAAAGTCAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 480

QY 481 CTGAGTTTGGAGTCCAGTCCAGCAAGCATCTGCTGTACAGAGATCAGCATGCTGCTCAAGA 540  
Db 481 CTGAGTTTGGAGTCCAGTCCAGCAAGCATCTGCTGTACAGAGATCAGCATGCTGCTCAAGA 540

QY 541 AGCAACAGCAACACAAAAGACACATCATAGACAGTGGTTCAGAAAGAGATCATCA 600  
Db 541 AGCAACAGCAACACAAAAGACACATCATAGACAGTGGTTCAGAAAGAGATCATCA 600

QY 601 GTAGTGAAGATTTGCGCGATGATCTGTCAGTGAAGCTTAAAGCAAGATTAAGTAAAGT 660  
Db 601 GTAGTGAAGATTTGCGCGATGATCTGTCAGTGAAGCTTAAAGCAAGATTAAGTAAAGT 660

QY 661 AAGACTGGCAAAATGTCAAGGTCAAGGTCAAGTATCAGCAAGTCTGATCTGATCTGATCTG 720  
Db 661 AAGACTGGCAAAATGTCAAGGTCAAGGTCAAGTATCAGCAAGTCTGATCTGATCTGATCTG 720

QY 721 CGGAGAGTGGGATTAAGAGAGTGTGAAGAAAGTGAATCTGACTATAGAGCAAAAAACA 780  
Db 721 CGGAGAGTGGGATTAAGAGAGTGTGAAGAAAGTGAATCTGACTATAGAGCAAAAAACA 780

QY 781 AAGTCAAAAGCCGTAAACCTCAAGCAGAAATTAAGCAAAAAGTGGGAAAAAGAGCAG 840  
Db 781 AAGTCAAAAGCCGTAAACCTCAAGCAGAAATTAAGCAAAAAGTGGGAAAAAGAGCAG 840

QY 841 GACGAGAGAGAGCAACTTGATTCATCAGAGAGAGAGAGAGATGATGAAGATTATG 900  
Db 841 GACGAGAGAGAGCAACTTGATTCATCAGAGAGAGAGAGAGATGATGAAGATTATG 900

QY 901 ATAGAGAGAGATCTGCTGCGCAGGCAACAGTGAATGTTAGTACAAAGAGCTGAAGAA 960  
Db 901 ATAGAGAGAGATCTGCTGCGCAGGCAACAGTGAATGTTAGTACAAAGAGCTGAAGAA 960

QY 961 CCAAGAGAGATTTGATGATTTGCTGGAAGTTTGTGAGAGAGATGCTCCACAGACTGAAG 1020  
Db 961 CCAAGAGAGATTTGATGATTTGCTGGAAGTTTGTGAGAGAGATGCTCCACAGACTGAAG 1020

QY 1021 AAGATGATTTGAACATATAGAGAAAGTTATGACACGTGCAATGGCCGGAAGAGAGCCA 1080  
Db 1021 AAGATGATTTGAACATATAGAGAAAGTTATGACACGTGCAATGGCCGGAAGAGAGCCA 1080

QY 1081 CTGGTGCCTCAACCAACATCTATGCGGTGAGGAGATGGGACCAAAATCTGGGTTTG 1140  
Db 1081 CTGGTGCCTCAACCAACATCTATGCGGTGAGGAGATGGGACCAAAATCTGGGTTTG 1140

QY 1141 AAAAGTCAAGAGAGCTGGGAGAAATACAGTATCTTATTAATGAAAGGCTGTACACA 1200  
Db 1141 AAAAGTCAAGAGAGCTGGGAGAAATACAGTATCTTATTAATGAAAGGCTGTACACA 1200

QY 1201 TCCATTAACCTTGGGAAACCTAAGAAAGCGTGAAGCAAAATGTTAAGGAATGACACA 1260  
Db 1201 TCCATTAACCTTGGGAAACCTAAGAAAGCGTGAAGCAAAATGTTAAGGAATGACACA 1260

QY 1261 AACTGCAACACTACAGAGAAAGAGATCAGAGACAAAACGCTGGCTGAAAAATGCTCTC 1320  
Db 1261 AACTGCAACACTACAGAGAAAGAGATCAGAGACAAAACGCTGGCTGAAAAATGCTCTC 1320

QY 1321 CAGAGATGTGGAATATTATTAATCTGCCAGAGAGAGCTTACAGATGATCTGCACAAACAT 1380  
Db 1321 CAGAGATGTGGAATATTATTAATCTGCCAGAGAGAGCTTACAGATGATCTGCACAAACAT 1380

QY 1381 ATCAAAATAGTGAAGAAATATTTGCTCATTCATCAAAAGTCAAGCTGCTGTTATCCGG 1440  
Db 1381 ATCAAAATAGTGAAGAAATATTTGCTCATTCATCAAAAGTCAAGCTGCTGTTATCCGG 1440

QY 1441 ACTACTATTTGCAAAATGCGAGGCTGCTGCTTACTGAGATGATGCTGGAGAGATGGTCTC 1500  
Db 1441 ACTACTATTTGCAAAATGCGAGGCTGCTGCTTACTGAGATGATGCTGGAGAGATGGTCTC 1500

QY 1501 TCATTGCGCAAAAGTTTCAGGACAGCATGTGATGATTTTATGAGCAAAATCATCCAGA 1560  
Db 1501 TCATTGCGCAAAAGTTTCAGGACAGCATGTGATGATTTTATGAGCAAAATCATCCAGA 1560

QY 1561 CTACTCCCTTTAAAGATCGCAAGGTTCTTAAAGCAGAACCAAGATTTGTTGCACTGAGA 1620  
Db 1561 CTACTCCCTTTAAAGATCGCAAGGTTCTTAAAGCAGAACCAAGATTTGTTGCACTGAGA 1620

QY 1621 AGCAACCATCTTACATTTGAGAGACATGAAGTCTGAGTTAAGAGATTATCAGTTAATG 1680  
Db 1621 AGCAACCATCTTACATTTGAGAGACATGAAGTCTGAGTTAAGAGATTATCAGTTAATG 1680

QY 1681 GATTGAATTTGGTGGCTGCTCATTCATGATGCAAAAGAAATGTTGATCTTGCAGATGAAA 1740  
Db 1681 GATTGAATTTGGTGGCTGCTCATTCATGATGCAAAAGAAATGTTGATCTTGCAGATGAAA 1740

QY 1741 TGGGTCTGGGTAAACAAATACAAATTTCTTTGCAACTACCTGTTTCATGAACATC 1800  
Db 1741 TGGGTCTGGGTAAACAAATACAAATTTCTTTGCAACTACCTGTTTCATGAACATC 1800

QY 1801 AACTGTATGGCCCTTTTCTCTGCGCGTGCACCTTCTTACCTTGTGACATCTTGGCAAGAG 1860  
Db 1801 AACTGTATGGCCCTTTTCTCTGCGCGTGCACCTTCTTACCTTGTGACATCTTGGCAAGAG 1860

QY 1861 AGATTCAAACTTGGGCTCTCCAGTGAATGCTGATGTTACTTAAAGGATATTAAGTAACTGA 1920  
Db 1861 AGATTCAAACTTGGGCTCTCCAGTGAATGCTGATGTTACTTAAAGGATATTAAGTAACTGA 1920

QY 1921 GAAATATGATTAAGACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
Db 1921 GAAATATGATTAAGACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980

QY 1981 TACTTCTGACGACATATGAATTTTACTGAAGATTAAGTATCTTGTGGTGTCTCAATT 2040  
Db 1981 TACTTCTGACGACATATGAATTTTACTGAAGATTAAGTATCTTGTGGTGTCTCAATT 2040

QY 2041 GGGCATTCATAGAGAGTTGATGAAGCTCATGCTTAAAAAATGATGACTCTTCTGTACA 2100  
Db 2041 GGGCATTCATAGAGAGTTGATGAAGCTCATGCTTAAAAAATGATGACTCTTCTGTACA 2100



Qy	2101	GGACCTTTAAATAGCTTTAAAGCCAAACATGCATCTCGATCTCTGAGAACCCCACTGCGAA	2160
Db	2101	GGACTTTAAATAGCTTTAAAGCTTAAAGCCAAACATGCATCTCGATCTCTGAGAACCCCACTGCGAA	2160
Qy	2161	ATTCCTCTCAAGAGCTGTGGCTCTTTCTTGATCTTCATCATGCCACAAAAATTTCTCTCT	2220
Db	2161	ATTCCTCTCAAGAGCTGTGGCTCTTTCTTGATCTTCATCATGCCACAAAAATTTCTCTCT	2220
Qy	2221	GGGAAAGATTTTGAGAGGAGCATGGCAAAGAGAGATGTGGTTATGCAAGCTTTCACA	2280
Db	2221	GGGAAAGATTTTGAGAGGAGCATGGCAAAGAGAGATGTGGTTATGCAAGCTTTCACA	2280
Qy	2281	AAGAGCTTTGAACCATTTTCTAATAAGAGAGTTAAAAAAGATGTAGAAAAGCTTTACGTG	2340
Db	2281	AAGAGCTTTGAACCATTTTCTAATAAGAGAGTTAAAAAAGATGTAGAAAAGCTTTACGTG	2340
Qy	2341	CTAAGGTTTGAACAAATTCGTAGAGATGGAATTAAGTCGATTCGACAGAACCATATTACAGT	2400
Db	2341	CTAAGGTTTGAACAAATTCGTAGAGATGGAATTAAGTCGATTCGACAGAACCATATTACAGT	2400
Qy	2401	GGATTTTAAACAGGAATTAATAAGCCCTCAGTAAAGGCTTCAAAAAGCGCAAGTACCTGAGCT	2460
Db	2401	GGATTTTAAACAGGAATTAATAAGCCCTCAGTAAAGGCTTCAAAAAGCGCAAGTACCTGAGCT	2460
Qy	2461	TTCTGACCATTTATGAGAACCTTAAAGATGTGTATACCATTTGCTACCTCATTAAGCCAC	2520
Db	2461	TTCTGACCATTTATGAGAACCTTAAAGATGTGTATACCATTTGCTACCTCATTAAGCCAC	2520
Qy	2521	CAGAGATTAATGAATTCATTAATAACAGAGGCTTACAGCATTTGATGATACGTAGAGAG	2580
Db	2521	CAGAGATTAATGAATTCATTAATAACAGAGGCTTACAGCATTTGATGATACGTAGAGAG	2580
Qy	2581	GGAAACCTAATCTCTTCTTGACACAGCTACGTATTCGTCTGCGAAGACGTGACACAGATTTC	2640
Db	2581	GGAAACCTAATCTCTTCTTGACACAGCTACGTATTCGTCTGCGAAGACGTGACACAGATTTC	2640
Qy	2641	TGATTTCTCTCAGATGTGTGAGAGATCTGAGACATCTAGCAAGATATCTGAGATATCGCC	2700
Db	2641	TGATTTCTCTCAGATGTGTGAGAGATCTGAGACATCTAGCAAGATATCTGAGATATCGCC	2700
Qy	2701	AGTTTCCCTTCAGAGACTTATGATGCATTAATAAAGGGAAATTGAGAAAGCAAGCACTGG	2760
Db	2701	AGTTTCCCTTCAGAGACTTATGATGCATTAATAAAGGGAAATTGAGAAAGCAAGCACTGG	2760
Qy	2761	ATCATTTTCAATGCAAGAGATCAGAGAGATTTCTGTTTTTACTGCTACCAAGAGCTGAG	2820
Db	2761	ATCATTTTCAATGCAAGAGATCAGAGAGATTTCTGTTTTTACTGCTACCAAGAGCTGAG	2820
Qy	2821	GATTTAGTATTAACCTTGCACTGCTGCAACACGTAGATATTTTATTCTAGCATGGAATC	2880
Db	2821	GATTTAGTATTAACCTTGCACTGCTGCAACACGTAGATATTTTATTCTAGCATGGAATC	2880
Qy	2881	CACCAATATGATTCGCGACAGCGACAGCGACAGCTATTAATATGAGACAGAAACACAGTTA	2940
Db	2881	CACCAATATGATTCGCGACAGCGACAGCGACAGCTATTAATATGAGACAGAAACACAGTTA	2940
Qy	2941	ATATTTTTCGGCTAGTACCAAAAGAGATCGATAGAAAGATTTCTGTAAGAGCCACAA	3000
Db	2941	ATATTTTTCGGCTAGTACCAAAAGAGATCGATAGAAAGATTTCTGTAAGAGCCACAA	3000
Qy	3001	AGAGATGTGTGCTAGACATTTAGTAAATTCAGAGATGCAACACAGAGAAAACTGTTTC	3060
Db	3001	AGAGATGTGTGCTAGACATTTAGTAAATTCAGAGATGCAACACAGAGAAAACTGTTTC	3060
Qy	3061	TGCATACAGGTTCACTGCATCAAGCTCTACACCTTTTAATAAGAGAGTTATACGCTA	3120
Db	3061	TGCATACAGGTTCACTGCATCAAGCTCTACACCTTTTAATAAGAGAGTTATACGCTA	3120
Qy	3121	TTTTGAAGTTTGGTGTCTAGAGAACCTTTTAAAGAACCTGAAGAGAGAAAGACAGAGCCCC	3180
Db	3121	TTTTGAAGTTTGGTGTCTAGAGAACCTTTTAAAGAACCTGAAGAGAGAAAGACAGAGCCCC	3180
Qy	3181	AGGAAATGGAATATNGATTAATAATCTTGAAGACAGCTGAAATCTGGGAAATATGACCAAGCTC	3240

Db	3181	AGGAATGATATAGTAAATCTTGAAAGAGCTGAAATGCGGAAATATBACCAAGTTC	3240
Oy	3241	CATTGACTGTAGGGGATGAGTTGCTTTCACAGTTCAAAAGTGGCACTTTTCCAAATATGG	3300
Db	3241	CATTGACTGTAGGGGATGAGTTGCTTTCACAGTTCAAAAGTGGCGAACTTTTCCAAATATGG	3300
Oy	3301	ATGAAGATGATATTGATTGGTGAACCAAGAAAGAAATTCAGAAATTTGGGAAGAAATCATCC	3360
Db	3301	ATGAAGATGATATTGATTGGTGAACCAAGAAAGAAATTCAGAAATTTGGGAAGAAATCATCC	3360
Oy	3361	CAGAAATCCAAACGAGAAAGGATAGAGAGAGAGAAACACAAAACAACTTGAAGAAATAT	3420
Db	3361	CAGAAATCCAAACGAGAAAGGATAGAGAGAGAGAAACACAAAACAACTTGAAGAAATAT	3420
Oy	3421	ACATCTCCCGAGAGATGAGAAACGTGCAAAAACAGATCACCTTATATGGGTGAAGGAA	3480
Db	3421	ACATCTCCCGAGAGATGAGAAACGTGCAAAAACAGATCACCTTATATGGGTGAAGGAA	3480
Oy	3481	GACGAGTAGAGACAGAGATATTCTGGATCTGATATGTGACCTCATCAGAAAGAAAC	3540
Db	3481	GACGAGTAGAGACAGAGATATTCTGGATCTGATATGTGACCTCATCAGAAAGAAAC	3540
Oy	3541	GGCCAAAAAAGCGTGGAGACCTCGAACCATCTCCGAGAGAAATTTAAAGAGTTATAG	3600
Db	3541	GGCCAAAAAAGCGTGGAGACCTCGAACCATCTCCGAGAGAAATTTAAAGAGTTATAG	3600
Oy	3601	ATGCAGAGATCAGGGGCTTTATCAAGATTAAGAAATTTGTGTGCGCTGTGGAAGGT	3660
Db	3601	ATGCAGAGATCAGGGGCTTTATCAAGATTTACAAAGATTAAGAAATTTGTGTGCGCTGTGGAAGGT	3660
Oy	3661	TGATATCTGTATAGTATGATCTGTAACTGGTGTATTAATGTGACACAGCTTAAGACGT	3720
Db	3661	TGATATCTGTATAGTATGATCTGTAACTGGTGTATTAATGTGACACAGCTTAAGACGT	3720
Oy	3721	TGGTGAACTTGTACATTAATGATGATTAAGCTTTTAAAGACAAATTCATCTGGACAAG	3780
Db	3721	TGGTGAACTTGTACATTAATGATGATTAAGCTTTTAAAGACAAATTCATCTGGACAAG	3780
Oy	3781	AAAGAGCAGAGAGGTAGACTTGGGAAAGTTTAAAGGCCCAAGCTTTGCAATCTCAGAGAGTGC	3840
Db	3781	AAAGAGCAGAGAGGTAGACTTGGGAAAGTTTAAAGGCCCAAGCTTTGCAATCTCAGAGAGTGC	3840
Oy	3841	AGGTAAATGCAAAACGTAGTCACTCTCACAGAAAGAAAGCTGGGACACCTGTGCACAAATCCA	3900
Db	3841	AGGTAAATGCAAAACGTAGTCACTCTCACAGAAAGAAAGCTGGGACACCTGTGCACAAATCCA	3900
Oy	3901	TTCTCTTCAGATCCAGAAAGAAAGAAAGATATGTATCCATCCATCCACACCAAGGCTGTC	3960
Db	3901	TTCTCTTCAGATCCAGAAAGAAAGAAAGATATGTATCCATCCATCCACACCAAGGCTGTC	3960
Oy	3961	ACCTCGATATATGATTTGGGTAAAGAGATGATCCATCTGTTAGTAGGCACTCATGATAT	4020
Db	3961	ACCTCGATATATGATTTGGGTAAAGAGATGATCCATCTGTTAGTAGGCACTCATGATAT	4020
Oy	4021	ATGGGTATGGCACTGGGAAATGATTAATAATGGATCCAGATCTCAGCTTAACACAGAGA	4080
Db	4021	ATGGGTATGGCACTGGGAAATGATTAATAATGGATCCAGATCTCAGCTTAACACAGAGA	4080
Oy	4081	TTTTTACCTGATATCCAGACAAAGAAACCCAGGCAAAAGCAGCTACAGACCCGTGCAGACT	4140
Db	4081	TTTTTACCTGATATCCAGACAAAGAAACCCAGGCAAAAGCAGCTACAGACCCGTGCAGACT	4140
Oy	4141	ACCTCATTAATTAATCTAGTAATAGACCTTGGCAAGAAAGACACAAAGGCTTGCTGTG	4200
Db	4141	ACCTCATTAATTAATCTAGTAATAGACCTTGGCAAGAAAGACACAAAGGCTTGCTGTG	4200
Oy	4201	CAGGCAATTCACAGAGAGAGACAAAGAAATTAAGAAATTAAGATGAAGGCTTCAAAA	4260
Db	4201	CAGGCAATTCACAGAGAGAGAGACAAAGAAATTAAGAAATTAAGATGAAGGCTTCAAAA	4260
Oy	4261	TAAAGAAAGAAATTAAGAGATATTCTTCACCAACCAACCTCAGAAAAATCTGATGAAGATG	4320

D	4261	TAAGAAGAAATTAAGAGATGATTTCTTCCACACCAACCTCAGAAAAATCTGATGAAGTG	4320
Q	4321	ATGAGAGGAGGAGTAACCAAGGTAAATGAATGAAATCTGAAAAATTAAGAAAAATCTAAAA	4380
D	4321	ATGAGAGGAGGAGTAACCAAGGTAAATGAATGAAATCTGAAAAATTAAGAAAAATCTAAAA	4380
Q	4381	AAATTCATTGCTGGATACTCCAGTTCATATTACTGCAACAGTAGAACAGTCCCTATCT	4440
D	4381	AAATTCATTGCTGGATACTCCAGTTCATATTACTGCAACAGTAGAACAGTCCCTATCT	4440
Q	4441	CAGAAAGATCTGAAAGAACTCCATCAGAAAGCAATTAGTGTGCAAAAGAAAGATGAGGC	4500
D	4441	CAGAAAGATCTGAAAGAACTCCATCAGAAAGCAATTAGTGTGCAAAAGAAAGATGAGGC	4500
Q	4501	CTGTCAAGCAGCAGCTGTAACACAGCTGGATAGACCAAGAGAGGGCCCTTCTGAAAGGAGC	4560
D	4501	CTGTCAAGCAGCAGCTGTAACACAGCTGGATAGACCAAGAGAGGGCCCTTCTGAAAGGAGC	4560
Q	4561	AGCTGGAACACTACTAGGCACTGCTTAATCAAAATTGGGGATCACATTACAGATGCTGA	4620
D	4561	AGCTGGAACACTACTAGGCACTGCTTAATCAAAATTGGGGATCACATTACAGATGCTGA	4620
Q	4621	AGGAGTACACAAATCCCGACCAATTAACACGGGAGGAAAAATTGGGATTTTGGT	4680
D	4621	AGGAGTACACAAATCCCGACCAATTAACACGGGAGGAAAAATTGGGATTTTGGT	4680
Q	4681	CCAAGTTTACAGATTGTGATGCCAGAAAGCTGCACAACTCTACAAACATCAATCAAAA	4740
D	4681	CCAAGTTTACAGATTGTGATGCCAGAAAGCTGCACAACTCTACAAACATCAATCAAAA	4740
Q	4741	AGGCCCAAGCTCTCAGCAACACATGACCAAAAACATTAGCAGCAATGTAAATACATG	4800
D	4741	AGGCCCAAGCTCTCAGCAACACATGACCAAAAACATTAGCAGCAATGTAAATACATG	4800
Q	4801	TAATCAGAAATCCAGATGTGAAAGACTGAAGAGCTCAAAACCATGATGATGTGTAGCA	4860
D	4801	TAATCAGAAATCCAGATGTGAAAGACTGAAGAGCTCAAAACCATGATGATGTGTAGCA	4860
Q	4861	GGGACAGTTATTTCTGTATGACATTTATCAAAATACCATGATCATCAAAAGCAGGC	4920
D	4861	GGGACAGTTATTTCTGTATGACATTTATCAAAATACCATGATCATCAAAAGCAGGC	4920
Q	4921	ATCAGGAGATGCTTACAGAAAAATGACTCCAGGAAAAAGCCATATTACGCTTCAGTA	4980
D	4921	ATCAGGAGATGCTTACAGAAAAATGACTCCAGGAAAAAGCCATATTACGCTTCAGTA	4980
Q	4981	ATGAAAGATGCACAGAGACTGGGACTACTACAACACGAGCAGCAGATACATACAGTATA	5040
D	4981	ATGAAAGATGCACAGAGACTGGGACTACTACAACACGAGCAGCAGATACATACAGTATA	5040
Q	5041	GTAACATAGAAAGTTAGATGACCAACAGGAGAGAGACACAGCTCAAACTGGAAGGAA	5100
D	5041	GTAACATAGAAAGTTAGATGACCAACAGGAGAGAGACACAGCTCAAACTGGAAGGAA	5100
Q	5101	ACTTAAAGACAGCCGGGGTCAATTCAGATACCGCTCCCATTCAGACCAGAGATACCT	5160
D	5101	ACTTAAAGACAGCCGGGGTCAATTCAGATACCGCTCCCATTCAGACCAGAGATACCT	5160
Q	5161	CAGATCACCGTCCACTTCAGATTAACGATCAATTAATCTTCGAGACATTAATGATACC	5220
D	5161	CAGATCACCGTCCACTTCAGATTAACGATCAATTAATCTTCGAGACATTAATGATACC	5220
Q	5221	ACTCAGACTGGCAAAATGGACACAGAGCTTCTGTAGTGGCCGAGGTACACACTAGATC	5280
D	5221	ACTCAGACTGGCAAAATGGACACAGAGCTTCTGTAGTGGCCGAGGTACACACTAGATC	5280
Q	5281	AGAGGTCTCCTTATGTTCAAGATCTCCCTAGGACACAGATCTCCATTTGAAACCTCAT	5340
D	5281	AGAGGTCTCCTTATGTTCAAGATCTCCCTAGGACACAGATCTCCATTTGAAACCTCAT	5340
Q	5341	CAGATCAAAAAAGTACACTGCAACATATCATGTGAATGACCGGAGACATTAACAAAGATGA	5400
D	5341	CAGATCAAAAAAGTACACTGCAACATATCATGTGAATGACCGGAGACATTAACAAAGATGA	5400

QY	5401	CATTTTCGACCTTCTTTTAGCCATATACAGTAACAAACAGTAATTCCTTCAT	5460
Db	5401	CATTTTCGAGACCTTCTTTTAGCCATATACAGTAACAAACAGTAATTCCTTCAT	5460
QY	5461	GACCTGAAGAAGATATGAGCTGGATATTCATATAGTAGAGCTTTCTTCCTTCCAGGA	5520
Db	5461	GACCTGAAGAAGATATGAGCTGGATATTCATATAGTAGAGCTTTCTTCCTTCCAGGA	5520
QY	5521	TGCAAGGTCATATATCCACAAGAAAGAAATATTTTGTATTTTAAAGTTATGCTGCAC	5580
Db	5521	TGCAAGGTCATATATCCACAAGAAAGAAATATTTTGTATTTTAAAGTTATGCTGCAC	5580
QY	5581	TGTGCTGCAANTGTGTGCGACCTTTTTTTAAAGAAATGGAAGATGTTTACTTTTACAGG	5640
Db	5581	TGTGCTGCAAAATGTGTGCGACCTTTTTTTAAAGAAATGGAAGATGTTTACTTTTACAGG	5640
QY	5641	GACCTCAACACTGCCCCCTTTCAGACTGGATCTTACTATAAACCTTCATATGCAAACTGG	5700-
Db	5641	GACCTCAACACTGCCCCCTTTCAGACTGGATCTTACTATAAACCTTCATATGCAAACTGG	5700
QY	5701	TTCTTAGGCTGGAACACAGATTAATTAATGTTTGTAAATGACACTTAAACACTGACCTGTG	5760
Db	5701	TTCTTAGGCTGGAACACAGATTAATTAATGTTTGTAAATGACACTTAAACACTGACCTGTG	5760
QY	5761	CTTATGTTTCAGGAAGAAGATGGGGATATTTATTTGTATTTATTTCTTGATAGAACCTGC	5820
Db	5761	CTTATGTTTCAGGAAGAAGATGGGGATATTTATTTGTATTTATTTCTTGATAGAACCTGC	5820
QY	5821	AAGGACTTGTCTACTTCCAAAGCTACTGTTTACATGTACACTGCGACACTTTCGC	5880
Db	5821	AAGGACTTGTCTACTTCCAAAGCTACTGTTTACATGTACACTGCGACACTTTCGC	5880
QY	5881	GCTTTTCATCACAAGCTTGATATTTAAATCTGTACTACAGTGTGAAAATAGCCAGGA	5940
Db	5881	GCTTTTCATCACAAGCTTGATATTTAAATCTGTACTACAGTGTGAAAATAGCCAGGA	5940
QY	5941	TTTTCTCTGTGTGATCAGTTATPATGCTTTTATGAAACAACAACAAACAAAAA	6000
Db	5941	TTTTCTCTGTGTGATCAGTTATPATGCTTTTATGAAACAACAACAAACAAAAA	6000
QY	6001	CAATTAATAAAAAAAAAACACAACAACAAACCAAAATGGCTGTAATTTTGTAAATTAAT	6060
Db	6001	CAATTAATAAAAAAAAAACACAACAACCAAAACCAAAATGGCTGTAATTTTGTAAATTAAT	6060
QY	6061	AAATAGGCTTTTTCCTCGACAGGCTTTTTCGGCTTCCTTCCCAACAACCTCAGGCT	6120
Db	6061	AAATAGGCTTTTTCCTCGACAGGCTTTTTCGGCTTCCTTCCCAACAACCTCAGGCT	6120
QY	6121	TCTTTTCACAAAGTCAGTATACTTACATGTTTAAATAAATATCTCGATGGAAATCAAAAT	6180
Db	6121	TCTTTTCACAAAGTCAGTATACTTACATGTTTAAATAAATATCTCGATGGAAATCAAAAT	6180
QY	6181	GTAATAAAGGGGAAGGGAATATTTATTCATCTTAAGTGCCTTTTATTTAGTACTTTT	6240
Db	6181	GTAATAAAGGGGAAGGGAATATTTATTCATCTTAAGTGCCTTTTATTTAGTACTTTT	6240
QY	6241	TACATACCTGTTTTGGTGTGTTATTTATTTTTTTTCTATTTAAACGTGCAGTGTG	6300
Db	6241	TACATACCTGTTTTGGTGTGTTATTTATTTTTTTTCTATTTAAACGTGCAGTGTG	6300
QY	6301	TGATGTTGTATAGAACAGTAGAATATCCACGCTAAAGCTGCGCCGGAAACCTTTTC	6360
Db	6301	TGATGTTGTATAGAACAGTAGAATATCCACGCTAAAGCTGCGCCGGAAACCTTTTC	6360
QY	6361	AGGTGCATTTGGTTTAAAAAGAGAAAGTGTCTATAGGTGAACACTTCAAAAACCAAGATCA	6420
Db	6361	AGGTGCATTTGGTTTAAAAAGAGAAAGTGTCTATAGGTGAACACTTCAAAAACCAAGATCA	6420
QY	6421	GCCAAAGATTCATGTAAATCCATTTGTTTCCCTCTTACATGGCAATPATGTCAAT	6480
Db	6421	GCCAAAGATTCATGTAAATCCATTTGTTTCCCTCTTAAATGCGGCAATPATGTCAAT	6480

[illegible]

QY	781	AAGTCAAAAGCCGTTAAACCTTC	CAAGCAGAAATTAAGCCAAAAAGTGGAAAAAGACACAG	840
Db	781	AAGTCAAAAGCCGTTAAACCTTC	CAAGCAGAAATTAAGCCAAAAAGTGGAAAAAGACACAG	840
QY	841	GACGAAAGAAGGACCACTTGATTCT	TCACGAGAGGAGGAGGAGCAGTACATGAAGTTATG	900
Db	841	GACGAAAGAAGGACCACTTGATTCT	TCACGAGAGGAGGAGGAGCAGTACATGAAGTTATG	900
QY	901	ATAAGAGAGGATCTCGTCGGCAGG	CAACAGTGAATGTTAGTTACAAAGACGTGAAGAA	960
Db	901	ATAAGAGAGGATCTCGTCGGCAGG	CAACAGTGAATGTTAGTTACAAAGACGTGAAGAA	960
QY	961	CCAGACAGANTCTGAGATTTCGTGG	AAGTTGTGTGAGAGAGTCTCCACAGACTGAAG	1020
Db	961	CCAGACAGANTCTGAGATTTCGTGG	AAGTTGTGTGAGAGAGTCTCCACAGACTGAAG	1020
QY	1021	AAGATGAATTTGAACATATAGACA	AGATTTATGACACGTGAATTTGGCCGAAAAAGAGCA	1080
Db	1021	AAGATGAATTTGAACATATAGACA	AGATTTATGACACGTGAATTTGGCCGAAAAAGAGCA	1080
QY	1081	CTGGTGCCCTCAACCACCATCTAT	TATGCCGTTGAGGACAGATGTTGACCCAAATGCTGGGTTTG	1140
Db	1081	CTGGTGCCCTCAACCACCATCTAT	TATGCCGTTGAGGACAGATGTTGACCCAAATGCTGGGTTTG	1140
QY	1141	AAAAGTCAAAAGGAGCTGGGAGAA	ATACAGATTTTATTAATGGAAGAGCTGTGCACCA	1200
Db	1141	AAAAGTCAAAAGGAGCTGGGAGAA	ATACAGATTTTATTAATGGAAGAGCTGTGCACCA	1200
QY	1201	TCCTATTAACCTTGGGAAACTGA	AGAAACGCTGACACACAAATGTTAAAGATGAACA	1260
Db	1201	TCCTATTAACCTTGGGAAACTGA	AGAAACGCTGACACACAAATGTTAAAGATGAACA	1260
QY	1261	AACGTGCAACTATACAAAGAAAA	AGATCAGAGACAACACGCTGGTGAAAAATGCTTCTC	1320
Db	1261	AACGTGCAACTATACAAAGAAAA	AGATCAGAGACAACACGCTGGTGAAAAATGCTTCTC	1320
QY	1321	CAGAAGATGTGGAATATTATTA	CTGCGCAGAGAGGCTTACAGATGATCGCAACAACAT	1380
Db	1321	CAGAAGATGTGGAATATTATTA	CTGCGCAGAGAGGCTTACAGATGATCGCAACAACAT	1380
QY	1381	ATCAAAATAGTGGAAAGAAATTA	ATGTCATTTCAATCAAAAGTCACACACTGGTTATCCGG	1440
Db	1381	ATCAAAATAGTGGAAAGAAATTA	ATGTCATTTCAATCAAAAGTCACACACTGGTTATCCGG	1440
QY	1441	ACTACTATTGCAAAATGCGACGG	CTCGCTTACTCAGAAATGATGCTGGGAGAGATGGTCTC	1500
Db	1441	ACTACTATTGCAAAATGCGACGG	CTCGCTTACTCAGAAATGATGCTGGGAGAGATGGTCTC	1500
QY	1501	TCATTGGCCAAAAGTTTCAGGAC	ACGATGTGATGATTTTATGCGAAGAAATCAATCCAGA	1560
Db	1501	TCATTGGCCAAAAGTTTCAGGAC	ACGATGTGATGATTTTATGCGAAGAAATCAATCCAGA	1560
QY	1561	CTACTCCCTTTAAGGACTGCGA	AGGTTCTTAAACACAGACCCAGATTTGTGTGCATGAGA	1620
Db	1561	CTACTCCCTTTAAGGACTGCGA	AGGTTCTTAAACACAGACCCAGATTTGTGTGCATGAGA	1620
QY	1621	AGCAACCATCTTACATTTGAG	AGACATGAAGATGCTGGAGTTAAGACATTAATCAGTTAAATG	1680
Db	1621	AGCAACCATCTTACATTTGAG	AGACATGAAGATGCTGGAGTTAAGACATTAATCAGTTAAATG	1680
QY	1681	GATGGAATTTGGCTCGCTCAT	TTCATGTGTCGAAGGAAATGTTGTATTTCTTGACAGTGA	1740
Db	1681	GATGGAATTTGGCTCGCTCAT	TTCATGTGTCGAAGGAAATGTTGTATTTCTTGACAGTGA	1740
QY	1741	TGGGCTTGGGTTAAACATATCA	AAACAATTTCTTTCTTGACACTACCTGTTTCATGAAATC	1800
Db	1741	TGGGCTTGGGTTAAACATATCA	AAACAATTTCTTTCTTGACACTACCTGTTTCATGAAATC	1800
QY	1801	AACGTATGATGCCCTTTTCTT	CTGCGCGGTGCACATTTCTACCTTGAACATCTTGGCAAGAG	1860
Db	1801	AACGTATGATGCCCTTTTCTT	CTGCGCGGTGCACATTTCTACCTTGAACATCTTGGCAAGAG	1860

QY	1861	AGATTCAAACTTGGGCTTCCTCAATGATGATGCTGTACTTTACTTTAGGAGATATTAACCTAGTA	1920
Db	1861	AGATTCAAACTTGGGCTTCCTCAATGATGATGCTGTACTTTACTTTAGGAGATATTAACCTAGTA	1920
QY	1921	GAATATGATATGAGACACTCATGANTGGATGCTACACAGACTTAAGCATTAAGTTAATACA	1980
Db	1921	GAATATGATATGAGACACTCATGANTGGATGCTACACAGACTTAAGCATTAAGTTAATACA	1980
QY	1981	TACTTTCGACACATATGAAATTTTACTGAGAGATTAAGTCATTTCTTGTTGGTGGTCTCAAT	2040
Db	1981	TACTTTCGACACATATGAAATTTTACTGAGAGATTAAGTCATTTCTTGTTGGTGGTCTCAAT	2040
QY	2041	GGGCAATTCATAGAGATTGAGAGCTCATGCTTTTAAAAATGATGACTCTTCTCTGTACA	2100
Db	2041	GGGCAATTCATAGAGATTGAGAGCTCATGCTTTTAAAAATGATGACTCTTCTCTGTACA	2100
QY	2101	GGACTTTAATATGACTTTTAATCTCAACATCGACTCTGTATTCGTGAAGCCCACTGCAAA	2160
Db	2101	GGACTTTAATATGACTTTTAATCTCAACATCGACTCTGTATTCGTGAAGCCCACTGCAAA	2160
QY	2161	ATTCCTCTCAAGAGCTGTGGTCTTTTGTGCATTTTCATCTGCGCAAAAAATTTTCTCTCT	2220
Db	2161	ATTCCTCTCAAGAGCTGTGGTCTTTTGTGCATTTTCATCTGCGCAAAAAATTTTCTCTCT	2220
QY	2221	GGGAAGATTTTTGAAGAGGACATGCGCAAAAGGAAGAGATGTGGTTATCGAAGTCTTACA	2280
Db	2221	GGGAAGATTTTTGAAGAGGACATGCGCAAAAGGAAGAGATGTGGTTATCGAAGTCTTACA	2280
QY	2281	AAGAGCTTGAACCAATTTTCTTAAGAAGATTAAAAAATGTATGAATAAGTCTTACTCTG	2340
Db	2281	AAGAGCTTGAACCAATTTTCTTAAGAAGATTAAAAAATGTATGAATAAGTCTTACTCTG	2340
QY	2341	CTAAGTGTGAACAAATTTCTGAGGATGGAATGAGATGTGCATTTGCAGAAGCAATATTACAAGT	2400
Db	2341	CTAAGTGTGAACAAATTTCTGAGGATGGAATGAGATGTGCATTTGCAGAAGCAATATTACAAGT	2400
QY	2401	GGATTTTAAACAAGCAATTTAATAAGCCCTCAGTAAAGGTTCAAAAGGCAGTACCTCAGGCT	2460
Db	2401	GGATTTTAAACAAGCAATTTAATAAGCCCTCAGTAAAGGTTCAAAAGGCAGTACCTCAGGCT	2460
QY	2461	TTTGCACATTTATGATGAGACTTAAAGAGGTTGTATACCATTTGCTACCTCATTTAAGCCAC	2520
Db	2461	TTTGCACATTTATGATGAGACTTAAAGAGGTTGTATACCATTTGCTACCTCATTTAAGCCAC	2520
QY	2521	CAGATGATTAATGAATTTCTATTAATTAACACAGAGGCTTACAGCATTTTGTATAGCTAGCAGC	2580
Db	2521	CAGATGATTAATGAATTTCTATTAATTAACACAGAGGCTTACAGCATTTTGTATAGCTAGCAGC	2580
QY	2581	GGAAACTATATCTTTCTTGAACAAGCTACTGATTCGTCTGCGAGACGCTGGCAACAGATTC	2640
Db	2581	GGAAACTATATCTTTCTTGAACAAGCTACTGATTCGTCTGCGAGACGCTGGCAACAGATTC	2640
QY	2641	TGATTTTCTCAGATGATGAGGATGCTGTGACACCTCAGACAAATATCTGGAAGTATCGCC	2700
Db	2641	TGATTTTCTCAGATGATGAGGATGCTGTGACACCTCAGACAAATATCTGGAAGTATCGCC	2700
QY	2701	AGTTTCCCTTCCAGAGACTTGTATGATCAATAAAGGGCAATTTGAGAGCAAGCACTGCG	2760
Db	2701	AGTTTCCCTTCCAGAGACTTGTATGATCAATAAAGGGCAATTTGAGAGCAAGCACTGCG	2760
QY	2761	ATCATTTCAATGACGAGAGATTCAGAGAGATTTCTGTTTTTACTGTCTCAAGAGCTGGAG	2820
Db	2761	ATCATTTCAATGACGAGAGATTCAGAGAGATTTCTGTTTTTACTGTCTCAAGAGCTGGAG	2820
QY	2821	GATTATGATTAATCTTGGCATCTGTCTGACACGTACTTATTTTGTATTTCTGACGTGAAATC	2880
Db	2821	GATTATGATTAATCTTGGCATCTGTCTGACACGTACTTATTTTGTATTTCTGACGTGAAATC	2880
QY	2881	CACAGAATGATCTGACAGGACACAGGCGAAGAGCTCATGATTTGGACAGAAAGCAAGGTTA	2940
Db	2881	CACAGAATGATCTGACAGGACACAGGCGAAGAGCTCATGATTTGGACAGAAAGCAAGGTTA	2940
QY	2941	ATATTATTCGGCTGTGTCACAAAGAGATCAAGTATGAGAGAGATATTTCTTAAAGAGCCAGAA	3000

```

|||||
Db 2941 ATATTATCGGCTAGTCAAAAAGATAGTAAGATATTTCTTGAAGACAAGA 3000
Oy 3001 ACAGAGATGCTAGACATTTAGTAATTCAGAGATGACAGAGAAAACCTGTC 3060
Db 3001 ACAGAGATGCTAGACATTTAGTAATTCAGAGATGACAGAGAAAACCTGTC 3060
Oy 3061 TGCATACAGGTTCACTCAACAGCTCTACACCTTTTAAAGAAAGGTTATACGTA 3120
Db 3061 TGCATACAGGTTCACTCAACAGCTCTACACCTTTTAAAGAAAGGTTATACGTA 3120
Oy 3121 TTTTGAAGTTGGTGTGTGAGAACTTTTAAAGAAAGGTTATACGTA 3180
Db 3121 TTTTGAAGTTGGTGTGTGAGAACTTTTAAAGAAAGGTTATACGTA 3180
Oy 3181 AGGAATGATATAGTGAATCTTGAAGAGCTGAACTCGGAGAAATGACGAGTC 3240
Db 3181 AGGAATGATATAGTGAATCTTGAAGAGCTGAACTCGGAGAAATGACGAGTC 3240
Oy 3241 CATTGACTGAGGGGATGATGTTGCTTCAAGTTCAGAGTGGCACTTTTCAATATG 3300
Db 3241 CATTGACTGAGGGGATGATGTTGCTTCAAGTTCAGAGTGGCACTTTTCAATATG 3300
Oy 3301 ATGAAGATGATATGATGATGAGACAGAAAAGAAATTCAGAAATTCAGATCC 3360
Db 3301 ATGAAGATGATATGATGATGAGACAGAAAAGAAATTCAGAAATTCAGATCC 3360
Oy 3361 CAGAAATCCACAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Db 3361 CAGAAATCCACAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Oy 3421 ACATGCTCCGAGAGATGAGAAATCTGCAAAACAGATAGCTTTAATGGAGTGAAGAA 3480
Db 3421 ACATGCTCCGAGAGATGAGAAATCTGCAAAACAGATAGCTTTAATGGAGTGAAGAA 3480
Oy 3481 GACGAGATGAGAGACAGAGATATTTCTGATCTGATAGATCCATCAGAGAGAGAG 3540
Db 3481 GACGAGATGAGAGACAGAGATATTTCTGATCTGATAGATCCATCAGAGAGAGAG 3540
Oy 3541 GGGCAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Db 3541 GGGCAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Oy 3601 ATGAGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 3601 ATGAGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Oy 3661 TTAGTCTGATAGATAGATGCTGATGATGATGATGATGATGATGATGATGATG 3720
Db 3661 TTAGTCTGATAGATAGATGCTGATGATGATGATGATGATGATGATGATGATG 3720
Oy 3721 TGGGTGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3780
Db 3721 TGGGTGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3780
Oy 3781 AAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3840
Db 3781 AAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3840
Oy 3841 AGGTGAATGCAAAAGTATGATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 3841 AGGTGAATGCAAAAGTATGATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Oy 3901 TTTCTTCAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 3901 TTTCTTCAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Oy 3961 ACTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4020
Db 3961 ACTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4020
Oy 4021 ATGGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4021 ATGGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080

|||||
Db 4021 ATGGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Oy 4081 TTTTACGTGATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 4081 TTTTACGTGATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Oy 4141 ACCGTATTAATTTCTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Db 4141 ACCGTATTAATTTCTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Oy 4201 CAGGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
Db 4201 CAGGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
Oy 4261 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
Db 4261 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
Oy 4321 ATGAGAGAGAGATACAGAG 4341
Db 4321 ATGAGAGAGAGATACAGAG 4341

RESULT 3
D14316 2292 bp mRNA linear VRT 03-FEB-1999
LOCUS delta-crystallin enhancer binding protein, complete sequence.
DEFINITION D14316
ACCESSION D14316
VERSION D14316.1 GI:391639
KEYWORDS ORF2.
SOURCE Gallus gallus (library: lambda gtl1) 13 day embryo lens cDNA to mRNA, clone Jf11.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2292)
AUTHORS Funahashi,J
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel. Aging and
Cancer, Tohoku Univ., 4-1 Seiryu-machi, Aoba-ku, Sendai 980-77,
Japan, Tel:022-272-9499, Fax:022-272-3982)
2 (bases 1 to 2292)
AUTHORS Funahashi,J, Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.
TITLE Delta-crystallin enhancer binding protein delta EPI is a zinc
finger-homeodomain protein implicated in postgastrulation
embryogenesis
JOURNAL Development 119 (2), 433-446 (1993)
MEDLINE 9411644
REFERENCE 3 (bases 1 to 2292)
AUTHORS Funahashi,J.
JOURNAL Unpublished (1994)
FEATURES
source
location/Qualifiers
1..2292
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="lens"
/clone_id="lambda gtl1"
/dev_stage="13 day embryo"
/note="Delta-gal fusion protein binds to blocks 10 to 3 of
HN fragment of delta-crystallin enhancer."
/codon_start=1
/product="ORF2"
/protein_id="BA03262.1"
/db_xref="GI:391640"
/translation="MELKKCNHCYLKPPDNEFYKOEALLOHLIRSSKLLILDKL
LIRLRNRNRYLIRFOSVYMLDLIAEYLKYROPFORDGSIKGLRKOALDFNVAEG
SEDFPCLSTRAGGLGINTASADIVYIEDSDMPOKDLOAORAHRIQKQKOVNRYL
VMGSEYEDILIERAKKMYLDHLYIORMDTTGVYHSGSPSSSPFNKEELASLTK
FGAEELFKPEPGEPOEDMDIDELIKRAETRENEPGLTVGDLLSPFYANFSNM
EDDILPEPNSRMPELIPESORPRITEPPEKXELLEYVMIPWRMCAKQISFNGCF
GRSRSRRTSGSDSDSTTERKPKRKGRPRITRENIKFSDBAIRFKISKKKGCF

```

LERIDAVARDAELVDKSEIDRLRIGELVHNCIKALKKNDKNSGDERAGRLGKYPTE  
RISGVQVNAKLVISSHEELAPLHKISIPSEDERRYRYPCHTKRAHNDIDWGEDSN  
LIVGIYEAJGSMCKWNGSRSQLNTEILDDPDRPROMSYRVPQTTSLNY"  
BASE COUNT 827 a 385 c 546 g 534 t  
ORIGIN

Query Match 21.3% Score 1410; DB 5; Length 2292;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1560; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2461 TTTCGACATATATGTAAGTAAGTGTGTAAACCATTTGCTACTCTTAAGCCAC 2520  
|||||  
DB 243 TTTCGAACATATATGTAAGTAAGTGTGTAAACCATTTGCTACTCTTAAGCCAC 302  
QY 2521 CAGATGATATGTAATTCATATATAACAGAGCCCTTACACATTTGATAGTACAGC 2580  
|||||  
DB 303 CAGATGATATGTAATTCATATATAACAGAGCCCTTACACATTTGATAGTACAGC 362  
QY 2581 GGAACATATCTCTTCTGACAGCTACTGATTCGTCTGCGAAGCGTGCAACAGAGTTG 2640  
|||||  
DB 363 GGAACATATCTCTTCTGACAGCTACTGATTCGTCTGCGAAGCGTGCAACAGAGTTG 422  
QY 2641 TGATTTCTCTCAGATGCTGAGATGCTGACATCCTGACAGATATCTGAATATGCGC 2700  
|||||  
DB 423 TGATTTCTCTCAGATGCTGAGATGCTGACATCCTGACAGATATCTGAATATGCGC 482  
QY 2701 AGTTCCCTCCAGAGACTTGATGATCAATAAAAGGGAATTGAGAGCAAGCAGCTG 2760  
483 AGTTCCCTCCAGAGACTTGATGATCAATAAAAGGGAATTGAGAGCAAGCAGCTG 542  
QY 2761 ATCATTTCAATGCAAGAGATCAAGAGATTTCTGTTTTTACTGTCTAACAAGCTGAG 2820  
543 ATCATTTCAATGCAAGAGATCAAGAGATTTCTGTTTTTACTGTCTAACAAGCTGAG 602  
QY 2821 GATTAGGATTAATCTGGCATCTGCTGACACTGATATTTGATTCCTACAGATC 2880  
DB 603 GATTAGGATTAATCTGGCATCTGCTGACACTGATATTTGATTCCTACAGATC 662  
QY 2881 CACAGAAATGATCTGAGGACAGGCGAGAGCTCATAGAAATTTGACAGAAAGAGTTA 2940  
663 CACAGAAATGATCTGAGGACAGGCGAGAGCTCATAGAAATTTGACAGAAAGAGTTA 722  
QY 2941 ATATTTCGCTTACTCAAAAAGAGATCAGTAGAAAGATTTCTGAAAGAGCCAGA 3000  
723 ATATTTCGCTTACTCAAAAAGAGATCAGTAGAAAGATTTCTGAAAGAGCCAGA 782  
QY 3001 AGAAGATGCTGTAGACCATTTAGTAATTCAGAGATGAGACAGAGAAACTGTTG 3060  
DB 783 AGAAGATGCTGTAGACCATTTAGTAATTCAGAGATGAGACAGAGAAACTGTTG 842  
QY 3061 TGCATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAATAAAGAGATTTACAGTA 3120  
843 TGCATACAGGTTCAACTCCATCAAGCTCTACACCTTTTAATAAAGAGATTTACAGTA 902  
QY 3121 TTTTGAAGTTGGTGTGAGAGACTCTTTAAAGAACTGAAAGAGAGACAGAGCCCC 3180  
903 TTTTGAAGTTGGTGTGAGAGACTCTTTAAAGAACTGAAAGAGAGAGAGAGCCCC 962  
QY 3181 AGGAATGATATGATGAATCTTTGAAGAGAGCGTGAACCGGGAATATGAGCCAGT 3240  
DB 963 AGGAATGATATGATGAATCTTTGAAGAGAGCGTGAACCGGGAATATGAGCCAGT 1022  
QY 3241 CATTGACTGTAGGGATGATGCTTTCACAGTTCAAGGTGGCAACTTTTCCATAATGG 3300  
DB 1023 CATTGACTGTAGGGATGATGCTTTCACAGTTCAAGGTGGCAACTTTTCCATAATGG 1082  
QY 3301 ATGAAGATGATATGATGAGTGGACAGAGAAATTCAGAAATTTGGAGAGAAATCATCC 3360  
DB 1083 ATGAAGATGATATGATGAGTGGACAGAGAAATTCAGAAATTTGGAGAGAAATCATCC 1142  
QY 3361 CAGAAATCCACGAGAGAGATAGAGAGAGAAAGCAAAAGAACTTAAAGAAATAT 3420  
|||||

DB 1143 CAGAAATCCCAAGAGAGAGATAGAGAGAGAGAAAGCAAAAGAACTTGAAGAAATAT 1202  
QY 3421 ACATGCTCCAGAGATGAGAAATCTGTCCAAACAGATCAGCTTTAATGGAGTGAAGAA 3480  
DB 1203 ACATGCTCCAGAGATGAGAAATCTGTCCAAACAGATCAGCTTTAATGGAGTGAAGAA 1262  
QY 3481 GAGCAGTAGAGAGAGAGATATTTCTGATCTGATAGTACTCCATCAGAGAAAGAAAC 3540  
DB 1263 GAGCAGTAGAGAGAGAGATATTTCTGATCTGATAGTACTCCATCAGAGAAAGAAAC 1322  
QY 3541 GGGCAAAAAGCGTGGAAAGACCTGAAACCATTCCTCCAGAGAAATATTAAGAGATTTGTG 3600  
DB 1323 GGGCAAAAAGCGTGGAAAGACCTGAAACCATTCCTCCAGAGAAATATTAAGAGATTTGTG 1382  
QY 3601 ATGCAGAGATGAGCGGTTTATCAAGAGTTTCAAGAAATTTGTGGCCCTCTGAGAAAGT 3660  
DB 1383 ATGCAGAGATGAGCGGTTTATCAAGAGTTTCAAGAAATTTGTGGCCCTCTGAGAAAGT 1442  
QY 3661 TAGATGCTGTAGTAGAGATGCTGAGATGTTGATTAATCTGAGACAGACCTTAGAGCTT 3720  
DB 1443 TAGATGCTGTAGTAGAGATGCTGAGATGTTGATTAATCTGAGACAGACCTTAGAGCTT 1502  
QY 3721 TGGGTGAACCTGTACATPATGATGATGCAATTAAGCTTTAAAGAGCAATTCATCTGAGAC 3780  
DB 1503 TGGGTGAACCTGTACATPATGATGATGCAATTAAGCTTTAAAGAGCAATTCATCTGAGAC 1562  
QY 3781 AAAGAGCAGAGATGAGACTTTGGAAAGTTAAAGCCCAACCTTTGATCTCAGAGATGTC 3840  
DB 1563 AAAGAGCAGAGATGAGACTTTGGAAAGTTAAAGCCCAACCTTTGATCTCAGAGATGTC 1622  
QY 3841 AGGTGAATGCAAACTACTCTCTCTCACGAGAAAGAGCTGGCACCTGACAAATTCGA 3900  
DB 1623 AGGTGAATGCAAACTACTCTCTCTCACGAGAAAGAGCTGGCACCTGACAAATTCGA 1682  
QY 3901 TTCCCTTCAGATCCAGAGAAAGAAAGATTTGATCATCCCATGCGCACACCAAGCTGTC 3960  
DB 1683 TTCCCTTCAGATCCAGAGAAAGAAAGATTTGATCATCCCATGCGCACACCAAGCTGTC 1742  
QY 3961 ACTTCGATATGATTTGGGTAAAGAGATGATTCATATCTGTTAGTGAAGCATATGAT 4020  
DB 1743 ACTTCGATATGATTTGGGTAAAGAGATGATTCATATCTGTTAGTGAAGCATATGAT 1802  
QY 4021 ATG 4023  
DB 1803 ATG 1805

RESULT 4  
AF006659 345 bp DNA linear VRT 30-OCT-1998  
LOCUS  
DEFINITION Gallus gallus CHD-2 (CHD-2) gene, partial cds.  
ACCESSION AF006659  
VERSION AF006659.1 GI:3811116  
KEYWORDS  
SOURCE  
ORGANISM chicken.  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 345)  
REFERENCE  
AUTHORS Griffiths,R., Double,M.C., Orr,K. and Dawson,R.J.  
TITLE A DNA test to sex most birds  
MOL. ECOL. 7 (8), 1071-1075 (1998)  
MEDLINE 98377745  
REFERENCE  
2 (bases 1 to 345)  
AUTHORS Griffiths,R., Double,M., Kate,O. and Dawson,W.  
TITLE Direct Submission  
SUBMITTED (02-JUN-1997) Zoology, Molecular Lab, Glasgow University,  
JOURNAL Glasgow G12 80Q, UK  
FEATURES  
location/Qualifiers  
1..345  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"

gene	CDS	BASE COUNT	Protein
	<pre> /chromosome="2" &lt;1..&gt;345 /gene="CHD-2" join(&lt;1..30,194..&gt;345) /gene="CHD-2" /codon_start=1 /product="CHD-2" /protein_id="AAC6644.1" /db_xref="GI:3811117" /translation="LPRRCRCAKQISFNGSGRRSRNSISGSDSDSITERRKPRKRR APRITPEENIKGFSPA" </pre>	<pre> 100 a      60 c      76 g      109 t </pre>	

Query Match	2.3%	Score 155;	DB 5;	Length 345;
Best Local Similarity	100.0%	Pred. No. 1.3e-66;		
Matches 155;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	3453	CAGATCAGCTTTAAAGGAGAGTGAAGACGACGACGTAGAGAGAGAAAGATTCTGGATCT	3512
Db	191	CAGATCAGCTTTAAAGGAGATGGAAGAACGACGATAGAGACGAAGAATTTCTGGATCT	250
Oy	3513	GATGTGATCCCTCATACAGAAAAGAAAGCGCAAAAAACGCGAAGAACTCGAACCAT	3572
Db	251	GATGTGATCCCTCATACAGAAAAGAAAGCGCAAAAAAGCGGAAAGCTCGAACCAT	310
Oy	3573	CCTCGAAGAAATTTAAAGATTTAGTGTGTCGACA	3607
Db	311	CCTCGAAGAAATTTAAAGATTTAGTGTGTCGACA	345

	RESULT	5			
	A58713				
	LOCUS	A58713	111 bp	DNA	linear PAT 06-MAR-1998
	DEFINITION	Sequence 32 from Patent W09639505.			
	ACCESSION	A58713			
	VERSION	A58713.1 GI:3714255			
	KEYWORDS	.			
SOURCE		unidentified.			
ORGANISM		unclassified			
REFERENCE		unclassified.			
AUTHORS		1 (bases 1 to 111)			
TITLE		Griffiths, R. and Tilvari, B.			
JOURNAL		AVIAN GND GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS			
COMMENT		Patent: WO 9639505-A 32 12-DEC-1996;			
FEATURES		ISLS INNOVATION (GB) Other publication AU 596996 961224. Location/Qualifiers			
source		1..111			
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT		43 a 20 c 25 g 23 t			
ORIGIN					

Query Match	1.7%	Score 111	DB 6	Length 111
Best Local Similarity	100.0%	Pred. No. 3e-44		
Matches 111	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	3498	AGATATTCGTGATCTGATAGTACCTCATCCACGAGAAACGCCAACCAAAAGCGTGTGA	3557	
Db	1	AGATATTCGTGATCTGATAGTACCTCATCCACGAGAAACGCCAACCAAAAGCGTGTGA	60	
Qy	3558	AGACCTCGAACCATTCCTCGAGGAAATATTTAAAGATTAGTGATCGAGAG	3608	
Db	61	AGACCTCGAACCATTCCTCGAGGAAATATTTAAAGATTAGTGATCGAGAG	111	
RESULT	6			
AF181828		2754 bp	mRNA	linear
LOCUS	AF181828			VFT 07-AUG-2001

DEFINITION Nymphicus hollandicus chromosome 2 chromodomin helicase DNA binding protein 1 (CHD1Z) mRNA, partial cds.

ACCESSION AF181828

VERSION AF181828.1

KEYWORDS GI:5917752

SOURCE cockatiel.

ORGANISM Nymphicus hollandicus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Psittaciformes; Cacoatidae; Nymphicus.

REFERENCE 1 (bases 1 to 2754)

AUTHORS Fridolfsson, A.K., and Ellegren, H.

TITLE Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes

JOURNAL Genetics 155 (4), 1903-1912 (2000)

MEDLINE 20384598

REFERENCE 2 (bases 1 to 2754)

AUTHORS Fridolfsson, A.-K., and Ellegren, H.

TITLE Direct Submission

JOURNAL Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 180, Uppsala S-75236, Sweden

FEATURES

Location/Qualifiers

1..2754

/product="chromodomains" name="histone H3" type="protein" id="AAB56025.1" /  
 /protein\_id="AAB56025.1" /  
 /db\_xref="gi:5917757" /  
 /translation="ELTDLRHKQOYIVRIIAHSNKSACGPRHCKNKGQSPVSECS  
 MEDLEAIARKEQARIDYEFSRNSQSTTFPFCKQVLAQORRVVALKQGYLIGHEHSE  
 LDVQLNGLFAMRLIASHMCKGNSCLADENGLAKTQITQISLVNLEHQLQYGLFVLP  
 LSTLSTMQREIQTVAASMANVAVYLDGITSRNHRIETHEHNNHQLKQTFNLTLYTEIL  
 LKDSFSGIOLNAPFAGVDEAHRLKNDSDLYKTLIDFKNHRLITDTPLOSLAKELM  
 SLTLEFIPKSFSEWGEFEEHSGKREYQVSAIKSELEPLRLRVKCKVSRSLPAVQO  
 ILIRNEMSLAQOYKQYKWLTPRYNVALSKSGKSGVSNLIMELKCKCNKCYSLKIPDD  
 NEFYNQDELQILIRSGKGLILDKMLRLIRLEGNVVLFGQAMVRLILAVLYKQ  
 PEPQDLSIKGELKQALNAGSDEDFELSTLRAGGILFMSADTVLIEFSQMN  
 NPQDLSIAQARAHRLGQKQVNIYLRKLVGEVEDILAEAKKQVLDHVLVLRMDTQY  
 KTVLHNGSPSSSTPFNKEELSLILFLGAEELRPEEGEGQPEMDIDELIKLRTFR  
 ENEGEPTLGVGDELLSQKVFANFSNDDDEDLIRPENCANNEELIPVQRRIIEEER  
 OKLEETLTVLPMRCKAQISFNGSEGRSSRSYSGSSQSIIEERKRRKGRGRTIT  
 PRENKGFSDAEIRFTISKYKSGFLRDLNVAQVLAQVLEISDILRGLGVYVNC  
 IKALKSSGGEORAGRLGKGVKQPTPIRSGVQNAVLVSHDESLDILKSPISPEEC  
 KRVRVLPCHTKAHDIDIMCKEDSDNSLLIGYEYGSWEIMIKMDPLSLTQKILIPDD  
 PDKK" /

Query Match	1.7%	Score 111:	DB 5:	Length 2754;
Best Local Similarity	100.0%:	Pred. No. 2,2e-44:		
Matches 111:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY	2352	CAAAATTCGAGCATGGAATGAGTCGATTCGACAGCAACATATTTCACAGTGGATTTTACA	2411	
Db	1000	CAAAATTCGAGCATGGAATGAGTCGATTCGACAGCAACATATTTCACAGTGGATTTTACA	1059	
QY	2412	AGCAATATTAAAGCCCTCAGTAAAGGTTCAAAAGGCGAGTACCTCAGGCTTT	2462	
Db	1060	AGCAATATTAAAGCCCTCAGTAAAGGTTCAAAAGGCGAGTACCTCAGGCTTT	1110	
RESULT	7			
LOCUS	A58664	153 bp	DNA	linear
				PAT 06-MAR-1998



DEFINITION	SEQUENCE 3 FROM PATENT WO9639505.
ACCESSION	A58684
VERSION	A58684.1
KEYWORDS	GI:3714247
SOURCE	unidentified.
ORGANISM	unidentified.
REFERENCE	unclassified.
AUTHORS	1 (bases 1 to 153)
TITLE	Griffiths, R. and Tlward, B.
	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
JOURNAL	BIRDS
COMMENT	PATENT: WO 9639505-A 3 12-DEC-1996;
FEATURES	ISIS INNOVATION (GB)
	Other publication AU 5906996 961224.
	Location/Qualifiers
source	1..153
	/organism="unidentified"
	/db_xref="taxon:32644"
BASE COUNT	58 a 40 c 31 g 24 t
ORIGIN	

Query Match	1.6%	Score 103	DB 6	Length 153	
Best Local Similarity	100.0%	Pred. No. 3.3e-40			
Matches 103	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	4100	CAAGAAACCCGACGCAAGACACCTACAGACCCGTCAGACTACTCATTAATTAATCTGAA	4159		
Db	51	CAAGAAACCCGACGCAAGACACCTACAGACCCGTCAGACTACTCATTAATTAATCTGAA	110		
QY	4160	TAAAGACCTTGCAAAAAGAAAGCAAAAGCTTGCTGTGCA	4202		
Db	111	TAAAGACCTTGCAAAAAGAAAGCAAAAGCTTGCTGTGCA	153		
RESULT	8				
LOCUS	AF181824				
DEFINITION	AF181824	2754 bp	mRNA	linear	VRF 07-AUG-2001
ACCESSION	Aegolius funereus chromosome W chromodomain helicase DNA binding protein 1 (CHD1W) mRNA, partial cds.				
VERSION	AF181824				
KEYWORDS	AF181824.1	GI:5917748			
SOURCE					
ORGANISM	Aegolius funereus.				
	Aegolius funereus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Archosauvia; Aves; Neognathae; Strigiformes; Strigidae; Aegolius.				
REFERENCE	1 (bases 1 to 2754)				
AUTHORS	Fridolfsson, A.K. and Ellegren, H.				
TITLE	Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes				
JOURNAL	Genetics 155 (4), 1903-1912 (2000)				
MEDLINE	20384598				
REFERENCE	2 (bases 1 to 2754)				
AUTHORS	Fridolfsson, A.-K. and Ellegren, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-AUG-1999) Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyavägen 18D, Uppsala S-75336, Sweden				
	location/Qualifiers				
FEATURES	1..2754				
SOURCE	/organism="Aegolius funereus"				
	/db_xref="taxon:103375"				
	/chromosome="W"				
	<1..>2754				
gene	/gene="CHD1W"				
	<1..>2754				
CDS	/codon_start=1				
	/product="chromodomain helicase DNA binding protein 1"				
	/protein_id="AA056021.1"				
	/db_xref="GI:5917753"				
	/translation="ELTIDDLKQYQYIERITIAHSNOKSASAGYPDYCKWQGLPYSECC				

BASE COUNT 943 a 461 c 623 g 727 t  
 ORIGIN

	Query Match	1.3%; Score 83; DB 5; Length 2754;
	Best Local Similarity	100.0%; Pred. No. 3 5e-30;
	Matches	83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	2919	ATTGCACAGAAAGAACAGGTTAATATTTTATCGGGCTAGTCACAAAAGATCAGTAGAAGAA 2918       ATTGGACAGAGAAAAACAGGTAAATTTTATCGGGCTAGTCACAAAAGATCAGTAGAAGAA 1626
Db	1567	
QY	2979	GATATCTTGAAAGAGCCACAGAA 3001       GATATCTTGAAAGAGCCACAGAA 1649
Db	1627	
RESULT	9	
AF006513		
LOCUS	AF006513	5947 bp mRNA linear PRI 27-NOV-1997
DEFINITION	Homo sapiens CHD1 mRNA, complete cds.	
ACCESSION	AF006513	
VERSION	AF006513.1	GI:2645428
KEYWORDS	.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 5947) Woodeage,T., Basrai,M.A., Baxevanis,A.D., Hieter,P. and Collins,F.S. Characterization of the CHD family of proteins Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)	
REFERENCE	97470991	
AUTHORS	2 (bases 1 to 5947) Woodeage,T.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, National Institutes of Health, 49 Convent Drive, Bethesda, MD 20892-4442, USA	
TITLE	Location/Qualifiers	
JOURNAL	1..5947	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="5"	
	/map="5q15-21; near WI-5811"	
	1..5947	
	/gene="CHD1"	
gene	164..5293	
CDS	/gene="CHD1"	
	/codon_start=1	
	/product="CHD1"	
	/protein_id="AA87381.1"	
	/db_xref="GI:2645429"	
	/translation="MNHSDEESRYRNSGSSOSDDSGASGSAGSSGSSGSSDGSS SOGSSSDSGSGSGSSGSSGSSSPENKRYOAKPKYDGAEPWKSSTPLAYORAIL KKOQQQQQQQHQAASSNGSEEDSSSEDSSSVKRRKHHDQMOMGSSGPSOS GSOSESSEERKSSCDTESDYEFKNVKNRKRPONRSKSNKKILQQRKRQIDSSSE	





QY 5836 TTTCAGAACTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAAG 5895  
|||||  
Db 133120 TTTCAGAACTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAAG 133179  
QY 5896 CTGGAATATTAAATTCGTAC 5917  
|||||  
Db 133180 CTGGAATATTAAATTCGTAC 133201

RESULT 12  
AC021449  
LOCUS  
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
pieces.  
AC021449 143079 bp DNA linear HTG 10-SEP-2000  
AC021449 3 GI:10047806  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
ORGANISM  
human.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
JOURNAL  
TITLE Homo sapiens, clone RP11-58M12  
REFERENCE  
2 (bases 1 to 143079)  
Unpublished  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
D'Amico, K., Dewar, K., Domino, M., Doyle, M., Fenesfor, J.,  
Fertella, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardana, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Leoczky, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
Macdonald, P., Margulis, N., McEwan, P., McGuck, A., McKenna, K.,  
McPheters, R., Melgrim, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Vtel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center clone name: L5154  
Center project name: L5154

Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 14400; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 38820: contig of 38820 bp in length  
38821 38920: gap of 100 bp  
38921 40411: contig of 1491 bp in length  
40412 40511: gap of 100 bp  
40512 43279: contig of 2768 bp in length  
43280 43379: gap of 100 bp  
43380 46905: contig of 3526 bp in length  
46906 47005: gap of 100 bp  
47006 51831: contig of 4825 bp in length  
51831 51930: gap of 100 bp  
51931 62619: contig of 10669 bp in length  
62620 62719: gap of 100 bp  
62720 75408: contig of 12689 bp in length  
75409 75508: gap of 100 bp  
75509 92516: contig of 17008 bp in length  
92517 92616: gap of 100 bp  
92617 106409: contig of 13793 bp in length  
106410 106509: gap of 100 bp  
106510 143079: contig of 36570 bp in length.

Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_lib="RP11-58M12 Human Male BAC"  
1. 38820  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"  
38921. 40411  
/note="assembly-fragment"  
40512. 43279  
/note="assembly-fragment"  
43380. 46905  
/note="assembly-fragment"  
47006. 51830  
/note="assembly-fragment"  
51931. 62619  
/note="assembly-fragment"  
62720. 75408  
/note="assembly-fragment"  
75509. 92516  
/note="assembly-fragment"  
92617. 106409  
/note="assembly-fragment"  
106510. 143079  
/note="assembly-fragment  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 1.2%: Score 82; DB 2; Length 143079;  
Best Local Similarity 100.0%; Pred. No. 7.7e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCAGAACTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAAG 5895  
|||||  
Db 133109 TTTCAGAACTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAAG 133108  
QY 5896 CTGGAATATTAAATTCGTAC 5917  
|||||  
Db 1331079 CTGGAATATTAAATTCGTAC 1331100

RESULT 13  
AC008531

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.  
AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 369535  
Center Clone Name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-coverage estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-coverage estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 56174: contig of 56174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 118190: contig of 4963 bp in length  
\* 118191 118280: gap of unknown length  
\* 118281 119684: contig of 1404 bp in length  
\* 119685 119784: gap of unknown length  
\* 119785 123597: contig of 3503 bp in length  
\* 123598 123599: gap of unknown length  
\* 123599 145659: contig of 22262 bp in length.  
FEATURES  
Location/Qualifiers  
1..145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-480B11"  
/clone\_lib="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 1.2%; Score 82; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 7.7e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCAGAGCTACTGTTTACATGTACACCTGGACACCTTGGCGCTTTTCACACAG 5895  
|||||  
Db 61620 TTTCAGAGCTACTGTTTACATGTACACCTGGACACCTTGGCGCTTTTCACACAG 61679  
|||||  
QY 5896 CTGGAATATTTAAATTCGTAC 5917  
|||||  
Db 61680 CTGGAATATTTAAATTCGTAC 61701  
|||||  
RESULT 14  
AC022121 219258 bp DNA linear PRI 30-AUG-2001  
LOCUS AC022121/c  
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
AC022121  
VERSION AC022121.6 GI:15375145  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced gi:15148108.  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence:  
Estimated total Number of Errors is 0.4.  
STS Content:  
WT-5811 G04974  
WT-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841.  
FEATURES  
Location/Qualifiers  
1..219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

Query Match 1.2%; Score 82; DB 9; Length 219258;  
Best Local Similarity 100.0%; Pred. No. 7.4e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCAGAGCTACTGTTTACATGTACACCTGGACACCTTGGCGCTTTTCACACAG 5895  
|||||  
Db 74578 TTTCAGAGCTACTGTTTACATGTACACCTGGACACCTTGGCGCTTTTCACACAG 74519  
|||||  
QY 5896 CTGGAATATTTAAATTCGTAC 5917  
|||||  
Db 74518 CTGGAATATTTAAATTCGTAC 74497  
|||||

## RESULT 15

AF181827

AF181827 2754 bp mRNA linear VRT 07-AUG-2001  
Nymphicus hollandicus chromosome W chromodomain helicase DNA

binding protein 1 (CHD1W) mRNA, partial cds.

ACCESSION

AF181827

VERSION

AF181827.1

KEYWORDS

GI:5917751

SOURCE

cockatiel.

ORGANISM

Nymphicus hollandicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Psittaciformes; Cacatuidae;

Nymphicus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

Search completed: August 3, 2002, 02:12:57  
Job time: 42547 sec

Mon Aug 5 11:51:35 2002

us-08-973-363-10.oli.rge

Page 15

**THIS PAGE BLANK (USPTO)**



XX Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determ. and to control sex of progeny  
XX  
PS Claim 1; Fig 5; 76pp; English.  
XX  
CC The chicken CHD-W gene (AAT42754) acting alone or in conjunction with  
CC the closely related CHD-1A gene (AAT42751) is suggested to initiate  
CC female development in birds. The sequence of CHD-1A was deduced  
CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
CC library using a great tit CHD-W sequence (see also AAT42755) as probe.  
CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
CC gene (see also AAT42736-37). It is located on an autosome or Z  
CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
CC specific signal on hybridisation to genomic DNA of a non-ferile  
CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
CC acids can also be used to control the sex of the progeny of a bird.  
XX  
SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 100.0%; Score 6608; DB 18; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCTCGGCAGCAGACGCCGCCGCCGACGCGAGCTCGGGCCGGGAAGGCTTGC 60  
1 cgggctcggcagcagacgc 60  
Db 1 cgggctcggcagcagacgc 60  
QY 61 CCGCCGAGCCGGAGAGCAGCAGCGATATTGGGCAAAAATCTTGCCATCTGAGAGATAG 120  
61 cccgccgagccggagagcagcagcgatattggggcaaaaatcttgccatctgagagatag 120  
Db 61 cccgccgagccggagagcagcagcgatattggggcaaaaatcttgccatctgagagatag 120  
QY 121 CAAGTCAACGCATTTACTTCGAAAACATACGAGTACCGAGAGAGGGATCTTGCCCTAC 180  
121 caagtcaacgcatttacttcgaaaacatacgagtagccgagagagggatcttgccctac 180  
Db 121 caagtcaacgcatttacttcgaaaacatacgagtagccgagagagggatcttgccctac 180  
QY 181 ACCTGTGTAACCTGAGTGGACTTCTTTTAACTTTAATCTTAATCAATGAATGGGCACA 240  
181 acctgtgtaacctgagtggaacttctttttaaactttaactttaacttaaatgagtcacaa 240  
Db 181 acctgtgtaacctgagtggaacttctttttaaactttaactttaacttaaatgagtcacaa 240  
QY 241 GTGATGAAGAAAGTGAAGAAACAGCACTGGAGAGTCAAGCAGATCAGATATGTTGTG 300  
241 gtgatgaagaaagtgaagaaacagcaactggagagtcagagcagatcagatgattgtg 300  
Db 241 gtgatgaagaaagtgaagaaacagcaactggagagtcagagcagatcagatgattgtg 300  
QY 301 GGTGACGTTGAGTTCGTGATCTGTTCAAGCTCTGGAAGCAGATGAGCATGGAATGACA 360  
301 ggtgacgttgagttcgtgatctgttcaagctctggaagcagatgagcatggaatgaca 360  
Db 301 ggtgacgttcagttcgtgatctgttcaagctctggaagcagatgagcatggaatgaca 360  
QY 361 GCCAGTCAGTACGATGACTCTGAATCTGTTGAGAGTCAAGCAGTCAATCCGATCAG 420  
361 gccagtcagtacgactctgaatctgttggagagtcagagtcgaatccgattcagatcag 420  
Db 361 gccagtcagtagcagtagcagtagcagtagcagtagcagtagcagtagcagtagcagtag 420  
QY 421 AGTGTGACACATCTAGAGAGAGAAACAGTTCAAGCTTAACCTCCGAAACCTCAGCGAT 480  
421 agtgtgacacatctagagagagaaacagttcaagcttaacctccgaaacctcagcgat 480  
Db 421 agtgtgacacatctagagagagaaacagttcaagcttaacctccgaaacctcagcgat 480  
QY 481 CTGAGTTTGGAGATCCAGTCCAGCATCTTGTGTACAGAGATCAAGCATGCTGTCAGA 540  
481 ctgagtttggagatccagttccagcatcttgtgtacagagatcaagcatgctgtcagaa 540  
Db 481 ctgagtttggagatccagttccagcatcttgtgtacagagatcaagcatgctgtcagaa 540  
QY 541 AGCAACAGCAACAGCAAAAAGCAGATCATACAGAGTGGTTCAGAGAGCATTCATCCA 600  
541 agcaacagcaacagcaaaaaagcagatcatacagagtggttcagagagcattcacatcca 600  
Db 541 agcaacagcaacagcaaaaaagcagatcatcagacagtggttcagagagcattcacatcca 600  
QY 601 GTAGTGAAGATTTCGGCATGATTCGTCCATGAACTAACTAAAGAAAAGATTAAGTG 660  
601 gtagtgaagatttcggcatgatttcgtccatgaaactaaactaaagaaaagatttaagt 660  
Db 601 gtagtgaagatttcggcatgatttcgtccatgaaactaaactaaagaaaagatttaagt 660  
QY 661 AAGACTGCAATGTACAGGTTCAGAGGTTCAGATCAGAACTGTTGATCTGAATCGG 720  
661 aagactgcaatgtacaggttcagaggttcagatcagaaactgttgatctgaatcgg 720  
Db 661 aagactgcaaatgtacaggttcagaggttcagatcagaaactgttgatctgaatcgg 720

QY 721 CGAAGATGGGGATMAAAGCAGTTGTGAAGAAAGTGAATCTGACTATGAGCCAAAACA 780  
721 cgaagatggggatmaaaagcagttgtgaagaaagtgaatctgactatgagccaaaaaca 780  
Db 721 cgaagatggggatmaaaagcagttgtgaagaaagtgaatctgactatgagccaaaaaca 780  
QY 781 AAGTCAAAAGCCGTTAAACCTCCCAAGCAGATTTAAGCCCAAAAAGTGGCAAAAGCCACAG 840  
781 aagtcaaaagccgtttaaacctcccaagcagatttaagcccaaaaagtggcaaaaagccacag 840  
Db 781 aagtcaaaagccgtttaaacctcccaagcagatttaagcccaaaaagtggcaaaaagccacag 840  
QY 841 GACGAAGAAAGCAACACTGATTCATCAGAGAGAGAGAGAGATGATTAAGATTATG 900  
841 gacgaagaaagcaaacactgatctcatcagagagagagagagatgatttaagattatg 900  
Db 841 gacgaagaaagcaaacactgatctcatcagagagagagagagatgatttaagattatg 900  
QY 901 ATAAGAGAGATCTGTCGCCAGGCAACAGTGAATGTTATTACAAAGAGCTGAAGAA 960  
901 ataagagagatctgtccgccaggcaaacagtgaattgtttacaaagagctgaagaa 960  
Db 901 ataagagagatctgtccgccaggcaaacagtgaattgtttacaaagagctgaagaa 960  
QY 961 CCAAGACAGATTTCTGATGATTTGCTGGAAGTTTGGAGAGAGATGTCACAGACTGAAG 1020  
961 ccaagacagatcttgatgatttctggaagtttggagagagatgtcacagactgaag 1020  
Db 961 ccaagacagatcttgatgatttctggaagtttggagagagatgtccacagactgaag 1020  
QY 1021 AAGATGAATTTGAACCTATAGAGAGTTTATGACAGTCAATGGCCGGAAGAGAGCA 1080  
1021 aagatgaatttgaacctatagagagtttatgacagtcaatggccggaaagagagca 1080  
Db 1021 aagatgaatttgaacctatagagagtttatgacagtctgaattggccgaaagagagca 1080  
QY 1081 CTGTCCTCAACCAACATCTATGCGCTTGAAGAGATGTTGACCCAAATGCTGGTTTG 1140  
1081 ctgtcctcaaccaacatctatgctgtgagagagatgttgacccaaatgctggtttg 1140  
Db 1081 ctgtgtcccaaccaacatctatgctgtgagagagatgttgacccaaatgctggtttg 1140  
QY 1141 AAAGTCAAAAGAGCTGGGAGAAATACGTAATCTTATTAATGGAAGAGCTGTGCACACA 1200  
1141 aaagtcaaaagagctgggagaaatacgtaattcttatttaaatggaaagctgtgcacaca 1200  
Db 1141 aaagtlcaaaagagctgggagaaatacagtagtattcttaataatggaaagctgtgcacaca 1200  
QY 1201 TCCATTAACACTTGGGAAACCTGAAGAAACGCTGAAGCAACAAAGTTTAAAGAAATGAACA 1260  
1201 tccattaacacttgggaaaactgaagaaacgctgaagcaaacaaagtttaagaaatgaaca 1260  
Db 1201 tccattaacacttgggaaaactgaagaaacgctgaagcaaacaaagtttaagaaatgaaca 1260  
QY 1261 AACTGGACACTACAGAGAAAAAGATCAGAGACAAAACGCTGCGTGAATAATCTTCTC 1320  
1261 aactggacactacagagaaaaagatcagagacaaaacgctgctgtaataatcttctc 1320  
Db 1261 aactggaacactacaagaaaaagatcagagacaaaacgctgctgtaataatcttctc 1320  
QY 1321 CAGAAAGATGGAATATATATACGTCCGACGAGAGCTTACGATGATCTGCACAAACAT 1380  
1321 cagaaagatgggaatatatatacgtccgacgagagcttaccgatgattctgcacaaacat 1380  
Db 1321 cagaagaatggaatatataacgtccagcagagagcttaccgatgattctgcacaaacat 1380  
QY 1381 ATCAAAATAGTGAAGAAATATATTCCTCATCAAAATCAAAATCAAGCAGCTGTTATCCGG 1440  
1381 atcaaaatagtgaagaaatatattcctcatcaaaatcaaaatcaagcagctgttattccgg 1440  
Db 1381 atcaaatatggaagaatatattgctcatcaaatcaaaatcaagcagctgttattccgg 1440  
QY 1441 ACTACTATTTGCAAAATGGCAGGGTCTGCCTTACTCAGAAATGTAGCTGGAGAGTGGTCTC 1500  
1441 actactatttgcaaaaatggcagggctgccttactcagaaatgtagctggagagatggtctc 1500  
Db 1441 actactatgcaaaatggaaggtctgccttactcagaaatgtagctggagagatggtctc 1500  
QY 1501 TCATTGGCAAAAAGTTTCAGCAGCATGATGATGATTTTGAAGAAATCAATCCACAGA 1560  
1501 tcattggcaaaaagtttcagcagcatgattgattgatttgaagaaatcaatccacaga 1560  
Db 1501 tcattggcaaaaagtttcagcagcatgattgattgatttgaagaaatcaatccacaga 1560  
QY 1561 CTACTCCCTTTAAGACTGCAAGGTTCTTAAACACAGACCAAGATTTGTTGCTACATGAGA 1620  
1561 ctactcccttttaagactgcaaggtttcttaaaacagagacaaagtttggctcagaga 1620  
Db 1561 ctactcccttttaagactgcaaggtttcttaaaacagagacaaagtttggctcagaga 1620  
QY 1621 AGCAACCATCTTCAATTGAGAGACATGAAGTCTGAGATTGAAGATTATAGTTAATG 1680  
1621 agcaaccatcttcaattgagagacatgaagtctgagattgaagattatagtttaagt 1680  
Db 1621 agcaaccatcttcaattgagagacatgaagtlccgaggttlaagattatcagttlaagt 1680  
QY 1681 GATTGAATTTGGCTGCTCATATCATGTCAGCAAGAAATAGTTGTAATCTTGCAATGAAA 1740  
1681 gattgaatttggctgctcatatcatgtcagcaagaaatagttgtaattcttgcaatgaaa 1740  
Db 1681 gattgaatttggctgctcatatcatgtgcaagaaatagttgtaattcttgcaatgaaa 1740  
QY 1741 TGGGCTGCGGTAATAACATACAAACAAATTTCTTTCTGAACTACCTGTTTCATGAATC 1800  
1741 tgggctgcggttaataacatatacaaaatcttcttctgaactactcgtttcaatgaatc 1800  
Db 1741 tgggtctggtgtaataacatacaaaatcttcttctgaactactcgtttcaatgaatc 1800  
QY 1801 AACTGTATGGCCCTTTTCTTCTGCGCGTGCCACTTTCTACCTTGACATCTTGGCAAAAGAG 1860  
1801 aactgtatggcccttttcttctgcgcggtgccactttctaccttgacatcttggcaaaagag 1860



```
Db 1801 aacgtatgagcccttctctccgctgcgcaccttccacttgcacatcttgcgaagag 1860
Oy 1861 AGATTTCAAACTGGGCTCCCTCAGATGATGCTGTGATTACTTAGAGATATTAACGTGA 1920
Db 1861 agattcaaaccttgggtccctccaga tgaatgctgtgttactttagagaataactagta 1920
Oy 1921 GAATATGATTAAGACTCATGAATGATGATCCAGACACTTAACATTTAAAGTTTAA 1980
Db 1921 gaaataatgaagaactcaatgaatgcatgcatccacagacaaagattaaagtttaca 1980
Oy 1981 TACTTCGACGACATATGAAATTTTACTGAAGATTAAGTATTCCTTGCTGCTCAATT 2040
Db 1981 tactcttgcagacataatgaatttacttgcgaagtaagttcaatcttcttgcgaatt 2040
Oy 2041 GGGCATTCATAGAGATGTTGATGAAGCTGTTTAAAAATGATGACTCTCTTCTGTA 2100
Db 2041 gggcatctcatagagtgatgtaagctcatcgttttaaaaaatgatagtctctctctgta 2100
Oy 2101 GGACTTTAATAGACTTTAAGTCACACATGACTTCTGATTAAGTCAGAACCCACATG 2160
Db 2101 ggaacttaataagctttaaagctccacacatgcgactcttgatctacggaaccacatgca 2160
Oy 2161 AATCCCTAAAGAGCTGTCGCTTGTGATTTGATTCATGATGCGAAAAATTTCTGCT 2220
Db 2161 attccctcaaaagctggtgctcttctgtgcatctcaatcgaacgaataatcttccctcc 2220
Oy 2221 GGGAAAGATTTTGAAGAGCATGCGCAAGGAAGAGATATGTTTATGCAACTTTCACA 2280
Db 2221 gggaaagattttgaagagcatgcgcaaggaagatgattgttatgcaacttttcaca 2280
Oy 2281 AAGACTTGAACCACTTTTACTAAGAAAGCTTAAAAAGATGTAGAAAAGCTTTTACCTG 2340
Db 2281 aagaaacttgaaccacttctacaaagaagttlaaaaaagatgtagaagaagcttctacctg 2340
Oy 2341 CTAAAGTTGAACAAATTTCTGAGATGGAATGAGTCATGTCAGAACAAATATTACAAGT 2400
Db 2341 ctaaagtttgaacaatctctgagga tgaataatgagtcattgcagaacaattlcaaat 2400
Oy 2401 GGATTTTAACAGAGAAATTTAAGCCCTCAGTAAAGCTTAAAGGACATACCTAGGCT 2460
Db 2401 ggaattttaacagaagaaattataaagccctcagtaaaggttcaaaagagttacctaaagct 2460
Oy 2461 TTCTGACATTTATGATGAACTTAAGAAGGTGTGAACATGCTGCTACCTATTAAGCCAC 2520
Db 2461 ttctgaaacttatgatacgaaacttaagaagtgtgtlaaccattgttaccctcataagccac 2520
Oy 2521 CAGATGATTAATGAATCTATTAATAACAGAGGCGCTTACACATTTGATGATAGCAG 2580
Db 2521 cagatgataatgaattctataataacagagagcttaccagcatttgatacgtagcagcg 2580
Oy 2581 GGAACCTAATCTTCTTGAACAAGCTACTGATTCGTCTGCGAAACGTGGCAACAGATGC 2640
Db 2581 ggaacctaaactctctcttgcgaagctactgacttcgctcgagaaagcttgcgaacagagttc 2640
Oy 2641 TGAATTTTCTCAGATGATGATGATGCTGACATCTCTAGCAAGAAATATCGAAGTTCGCC 2700
Db 2641 tgaaatttctctccagatggtgagagatgctgcgaactccctcagaaatcatctgaagatacgc 2700
Oy 2701 AGTTTCCCTTCAGAGACTTGATGATCAATAAAGGGAAATTGAGGAAGCAACACTGG 2760
Db 2701 agtttcccttccagaaacttgatgatcaataaagggaatttgaaggcaagaactgg 2760
Oy 2761 ATCATTTCAATGACAGAGATCAGAGATTTCTGTTTTTACTGCTTACAGAGCTGGAG 2820
Db 2761 atcaattcaatgaagaagatcagaagattctggttttcttacttgcataaagaagctgag 2820
Oy 2821 GATTAGTATTAACTTGGCATCTGCTACACTGTAGTTATTTTGGATTGCTGACTGAATC 2880
Db 2821 gattaggtattaaacttgcgcatctgcgacacgttagtatttgcattctgcagctgaaatc 2880
Oy 2881 CACAGAAATGATTCGACAGCAGCGAGAGCTCATGGAATTTGACAGAGAAACAGGTTA 2940
Db 2881 cacaagaatgatctgcgagcagcgcgagagctcatagaatttgcacagaagaacagttta 2940
Oy 2941 ATATTTATCGGCTTATGTCACAAAAAGGATCAGTAGAAGACATATTTCTTGAAGGCCAAGA 3000
Db 2941 atattatcggcttagtctacaaaaggatcagtagaagaataattcttgcgaagaagccaaaga 3000
Oy 3001 AGAAGATGCTGCTAGACATTTTATGTAATTCAGAGAAATGAGACACGAGAAAAATCTGTC 3060
Db 3001 agaagatgctgttagacacatttagtaattcagagaa tgcacgcagacagaaactgttcc 3060
Oy 3061 TGCATPACAGGTTCACTCCATCAACCTCTACACTTTTATTAAGAGATTTATCAGCTA 3120
Db 3061 tgcatacaggttcaactccatcaacgaagctctacacctttaaataaagaagttatcacgtta 3120
Oy 3121 TTTTGAAGTTTGTGCTGAGAGACTCTTTAAAGACCTGAAGGAGAAAGAGAGGCCCC 3180
Db 3121 ttttgaagtttgtgtctgaggaactctttaagaacctttaagaaggaagaaacagggcccc 3180
Oy 3181 AGGAAATGATATATGATGAAATCTTGAAGAGACTGAAACCTCGGAAAAATGAGCCAGTCC 3240
Db 3181 aggaataatgatatagatgaatacttgaaagagagcttgaabctcgggaaaaatgagccaagttc 3240
Oy 3241 CATGACATGTAAGGAGATGAGAACTCTTTCACAGTCAAGGTGCGAACTTTTCCATATGC 3300
Db 3241 cattgacctgaagggagtgagtgcttccacagttcagaagtgcgaaccttcccaataatgg 3300
Oy 3301 ATGAGATGATATTTGACTTGGAAACCAAGAAATTTCAAGAAATTTGGAGAAATTCATCC 3360
Db 3301 atgaagaatgatattgacttggaaaccaagaaatttcaagaaatttggagaaatctcatcc 3360
Oy 3361 CAGATCCCAAGGAGAGATAGAGAGAGAAAGACAAAGAAATCTTGAAGAAATAT 3420
Db 3361 cagaatcccaacggagaaagatcagaagggaggaagaagacaagaagaaacttgaagaataat 3420
Oy 3421 ACATGCTCCGAGGATGAGAACTGTGCAAAAACAGATCAGCTTTAATGGAGTGAAGGAA 3480
Db 3421 acatgctcccgaggatgagaataacttgcataaacagatcagctttaaaggagtgaaaggaa 3480
Oy 3481 GAGCAGTAGAGACGAGAAGATATTTCTGATCTGATAGTACTCCATCAGGAAAGAAAC 3540
Db 3481 gagcagtagagagagaagaatattctgatactgatagttactccatacagagaagaagaac 3540
Oy 3541 GGCCAAAAAAGCGTGAAGACCTCGAACCATCTCTCGAAGAAATTTAAAGATTATAGT 3600
Db 3541 ggcctaaaaagcgtggaagagcttcgaaccatctctcggaaataattaaagatttagtg 3600
Oy 3601 ATCCAGATCAGCGGCTTTATCAAGAGTTTCAAGAAATTTGTGGCCCTCTGGAAGAGT 3660
Db 3601 atccagatcagcggtttatcaagaagttacaagaatlttgggtccctctggaagaggt 3660
Oy 3661 TAGATCTGTAGCTAAGAGATGCTGAACGTGTGATTAATCTGAGACAGACTTGAAGCT 3720
Db 3661 tagatctgtatgcagaagatgctgaactggttgaataaacttgcagacgaacttaagaagtt 3720
Oy 3721 TGGGTAACTTGTATCAATATGATGATCAATTAAGGCTTTAAAGACAAATTTCTGAGCAAG 3780
Db 3721 tgggtaaacttgtatcaatataatgataatgataatgataatgataatgataatgataatg 3780
Oy 3781 AAAAGCAGAGAGTGAAGCTTGGGAAAGTTAAAGCCCAAGCTTTCAGATCTCAGAGTGC 3840
Db 3781 aaagagcagaggttagacttgggaaagttlaaaggcccaagcttctgcgaatctcgaagtg 3840
Oy 3841 AGGTGAATGCAAAACTATGTCATCTCAGAAAGAGAGCTGGCACCACTGCACAATCCA 3900
Db 3841 aggtgaatgcaaaaactatgatactctcagaaagaagcttgcacacacttgcacaatacca 3900
Oy 3901 TTCTCTCAGATCCAGAAAGAAAGAAAGATATGTCATCCATGCCACACCAAGCTGCTC 3960
Db 3901 ttctctcagatcccaagaagaagaagataatgtcatcccatgccaacacaagaagctgctc 3960
Oy 3961 ACTTCGATATGATTGGGTTAAAGAAAGATGATTCGAATCTGTTAGTGGCATATATGAT 4020
Db 3961 acttcgatatagattgggtlaaagaagatgattcccaactgtttagtgatgacatcatgaa 4020
```

QY 4021 ATGCGTATGCGAGCTGGGAAATGATTAATAATGATTCAGATCTCAGCTTACACAGAGA 4080  
|||||  
Db 4021 atgctatgagcgttggaaatgataaaaaatgataccgactctcaacacacagaga 4080  
QY 4081 TTTTACCTGATGATTCACAGACAGAAACCCAGGCAAAAGCAGCTACAGACCCGTCAGACT 4140  
|||||  
Db 4081 tttaacctgatacgcagacaagaacccagcaaaagcagctacaagcccgctgagact 4140  
QY 4141 ACCCATTAATTAATCTGATTAAGACCTTGCAAGAAAGAGCAACAAAGCTTCTGCTGTG 4200  
|||||  
Db 4141 acccatlaaatctacgaaataaagaccttgcaagaagaagcacaagagcttgcgtg 4200  
QY 4201 CAGGCAATTCACAGAGAGAGAGACAGAAATAGAGAAATAGATGAAGCTTCAAAAA 4260  
|||||  
Db 4201 caggcaatctcaagagagaagagacaagaataaagaataaagatgaaagcttcaaaaa 4260  
QY 4261 TAAAAAGAAAGATTAAGAGTATTTCTTCACCACAAACCTCAGAAAATCTGATGAAGATG 4320  
|||||  
Db 4261 taaaagaagaataaagatgattcttccacacaaaccttcagaaaaatctgataagatg 4320  
QY 4321 ATGAGAGAGGATACAGATTAAGTAAATGAAATCGAAATTAAGAAATCTTAA 4380  
|||||  
Db 4321 atgagagaggaagatacaagaatgaaatgaaatctgaaatacaagaataatctlaaa 4380  
QY 4381 AAATTCATTCGCTGGATCTCAGTACCTCATATTACTGCAACCCAGTGAACCAATTCCTATCT 4440  
|||||  
Db 4381 aaatccatctgctggtatctcagcttcaatctactgcaaccagtgaaaccttccatct 4440  
QY 4441 CAGAGAAATCTGAAAGAACTCCATACAGAAAGACATTAGTGTGCAAGAAAGATGAGGC 4500  
|||||  
Db 4441 cagaagaatctgaaagaaactccatcagaagaatctagtgctgcaagaagaatgaggc 4500  
QY 4501 CTGCAAGAGCAGCAGCTGAAACAGCTGTGATAGCAGAGAAAGGCTTCTGAAAGGAGC 4560  
|||||  
Db 4501 ctgcaaaagcagcagctgaaacagctgataagacaagaagaagccttctgaaagagac 4560  
QY 4561 AGCTGGAACATCTAGGCACTGTCTTAATCAAAATTTGGGATCAGATTACAGAAATGCCGA 4620  
|||||  
Db 4561 agctggaacatactagcagtgcttaatacaaaatlgggatcacatcagaatctgcga 4620  
QY 4621 AGGAGTACACAAATCCGAGCAATTAATAACAGTGGAGAGAAAAATTTGTGGATTTTGTGT 4680  
|||||  
Db 4621 agggatcacaaaatcccgagaaaataaaacagtgagagaaaatctgtggttcttgg 4680  
QY 4681 CCAAGTTTACGAATTTGATGCCAGAAAGCTGCACAAACTGTACAAACATGCAATCAAAA 4740  
|||||  
Db 4681 ccaagtttaagaatttgatgagcagaagaagctgcacaaactctacaacatgcaatcaaa 4740  
QY 4741 AGGCGCAAGAGCTTCAGCACACACAATGACCAAAACATTAGCAGCAATGTGAATACATG 4800  
|||||  
Db 4741 agcgccaagagcttcagacaacaatlgacaaaacattagcagaatctgatacacatg 4800  
QY 4801 TAAATCAGAAATCCAGATGTGGAAGAGCTGAAAGAGACTACAAACATGATATGTAAGCA 4860  
|||||  
Db 4801 taaatcaaaaatccagaatgtaagaagactgaaagagactaaacacatgatagtatgagca 4860  
QY 4861 GGGACAGTTATTTCTTGATAGACATTATTCACATACCATGATCATCAGAAAGCAGC 4920  
|||||  
Db 4861 gggacagttaattctctcgtatgatacattatcaatacagatcatalcaacaagaagc 4920  
QY 4921 ATCAGGAGATGCTTACAGAAAAAGTACTCAGGAAAAGGCCATATTACGCTTCACTA 4980  
|||||  
Db 4921 atcagggatgacttacaagaagaatgactcaggaagaagccatatctcagccttcagta 4980  
QY 4981 ATGGAAGAAAGATCAGAGAGACTGGGATCATCTCAAAACAGAGAGAGATGACTACAGTATA 5040  
|||||  
Db 4981 atggaaaaagatcacagagacccgggatacactacaacacagagatacactacagtgata 5040  
QY 5041 GTAACATATGAAAGTTAGATGATCAGACAGAGACAGACAGCTGAAACCTGGAAGGAA 5100  
|||||  
Db 5041 gtaaacatagaaagttagatgataccacagagcgagagaccataggttcaaaccttgaaagaa 5100

QY 5101 ACTTAAAAAGACGCCGGGATCATTCAGATCACCCGCTCCATTTAGACACCAGATACACT 5160  
|||||  
Db 5101 acttaaaaaagacgccggggtatcttcagataccacgctcccatctcagaccacagataacc 5160  
QY 5161 CAGATCACCGTTCCACTTCAGAAATACAGCCATCATTAATCTTGAGAGATTATAGATACC 5220  
|||||  
Db 5161 cagataccgcttcacatcagataacagccaataaactctcagagattatagataacc 5220  
QY 5221 ACTGAGCTGGCAATGGACACAGAGCTTCTGATGAGGCGCCGAGGTCACCACTACATAC 5280  
|||||  
Db 5221 accagactggcaaatggacaacagagcttctggaagagcgcgcgaaggtcacactagac 5280  
QY 5281 AGAGGTCTCTTAATGATTCAGATCTCCCTAGACACAGATCTCCATTTGAACACTCAT 5340  
|||||  
Db 5281 agaggtctcttaattgattcaagatctccctcaggacaagcttccatcttgaacacatc 5340  
QY 5341 CAGATCACAAAGTACACTTACATACATAGTGAATAGCCGAGACATTAACAAAGACTGA 5400  
|||||  
Db 5341 cagatacaaaaagtacacttgaaacataatggaatagccggaaagacatacaaaagactga 5400  
QY 5401 CATTTTCGACCTTTCTTTTAAAGCCATATACATGAATTAACACAGTAATTGCCCTTACAT 5460  
|||||  
Db 5401 catlttcggacctcttctttagccaatacagatacaactaaacagataatctgcttaacat 5460  
QY 5461 GACTTGAAGAATGACTGATATTTCTATACATAGCAGATATGCTTACTTCTTCCAGA 5520  
|||||  
Db 5461 gacttgaagaatgactgataatcttatcatalcagtagcagatgttactcttccagga 5520  
QY 5521 TGCAAGGTCTTATTTCCACACAGAAAGAAATATTTTGTATTTAAAGTTATCTGTCAC 5580  
|||||  
Db 5521 tgcgaaggtcttatttccacacagaaagaaaatcttctgataatgaatgatactgcac 5580  
QY 5581 TGCTGTCAAAATGTTGTGACCTTTTAAAGAAATGGAAGATGTTTACTTTTAAAGC 5640  
|||||  
Db 5581 tgctgtcaaaatgttgtgacctttttaaagaaatggaagatgattacttcttaagc 5640  
QY 5641 GACCTCAACACTGCCCCCTTTCAGACTGATCTACTTAAACCTTCTATGTCAAAGTG 5700  
|||||  
Db 5641 gacctcaaacactgcccccttcaagactgactcttctaataaaccttcaatctgcaaa 5700  
QY 5701 TTCTAGGCTGACACAGATTAATTAATGTTTGTAAATGAACACTTAAACCTGACCTGTG 5760  
|||||  
Db 5701 ttctagagctgaaacagatlaaatlaactgctgtaaaagaaactlaaacctgcgcgtg 5760  
QY 5761 CTATGTTTTCAGAAAGATGGGGATTTATTTGTTTATTTCTTGTGAGAACTCTC 5820  
|||||  
Db 5761 ctatgcttcaagaaagaaatgaggtatttattgttcttcttcttgtagagacctc 5820  
QY 5821 AAGGACTTTGTTCACCTTCCAAAGCTACTTGTATGATGTGACCTGACACCTTGCC 5880  
|||||  
Db 5821 aaggacttgttcaacttccaaagctactgttctacatgtaacactgcgacacacttgc 5880  
QY 5881 GCTTTTATCACAAGTGTGAATATTTAAATTTCTTACTCTACGTTGTAATAATGCCGGA 5940  
|||||  
Db 5881 gcttttataccaagcttgaaatlaataatctgtaactaaagcttgtaaaatgaccgga 5940  
QY 5941 TTCTCCGTGTTTGTAGATGATTAATGCTTTTATGAAACAAACAAACAAACAAAA 6000  
|||||  
Db 5941 ttctccgttgttgaatcagatlaataatgcttttattgaaacaaaacaaaacaaaaaa 6000  
QY 6001 CAATTTAAAAAAAACACAAACAAACCAACAAATGCTGTAAATTTTGTAAATTAAT 6060  
|||||  
Db 6001 caatttaaaaaaaaacacaaacaaacaaacaaatgctgtaaattatctgtaaatlaatt 6060  
QY 6061 AAATGAGCTTTTTCGTCAGGCTTTTGGCTTGGCTTCCCAACAACTCAGGCT 6120  
|||||  
Db 6061 aaatgagcttcttccgtcagagcttcttctgagcttccctcccaacactcagagct 6120  
QY 6121 TCTTTTCAAAAGTCACTATCTATACATGTTTATTAATAATATCTCGATGGAATCAAAAT 6180  
|||||  
Db 6121 tctttcaaaaagtcagatatacttaacatgcttataataaatatctcgatgatacgaat 6180  
QY 6181 GTAAAAATGGGGAAGGAATATTTTATTCATTTTAAAGTGTCTCTTTTATTTGATACTTT 6240

Df	6181	gtaaaaatgg99aaaggaatactttattccatttagtgcctcttttatlttgatacctt	6240
Oy	6241	TACATACAGTTTTTGGGTGCTTTATTATTTTATTTTTCATATAAAGTCAGTGTTG	6300
Df	6241	tacataacctgtcttgggtgttgtttattattatttttttttcattaactgcgtacgtgtg	6300
Oy	6301	TGATTGTTGAATGAACAGTAGAATATCCCACTTAACCTGTGCCCGGAAGAAGCTTTC	6360
Df	6301	tgatttgttgtaataagaacagtgagaaatatoccatctcaaacgtgtgccctggaaagctttc	6360
Oy	6361	AGGTCATTGGTTTAAAGAAAGCAACTGCTCTATAGTGAACAATCAACCAAGACCA	6420
Df	6361	agggtgacttggctttaaaaaaggaagtgcttcataagtgacaaccataaaccagatca	6420
Oy	6421	GCCAGATTCACTTGAATCCACTTTGTTTCCCTCTTTAACATGGGCATTAATGTCAAT	6480
Df	6421	gccaaagtccaatctgtaaaccatctgtttcccctcttaacatgvgcacaatgtaaat	6480
Oy	6481	GTGCTATGACAGCATTAATATTTTACAGATTTGAATGACTTTATTAAAGAAATGTAC	6540
Df	6481	gtgcctaagcagaactaataatttaagaagattgaaagactttaacagaaattgttac	6540
Oy	6541	AATGCACACTGATTTGACATAGATTAACCTTCTATCTGACAAATTTAAATTAACATAAACCNA	6600
Df	6541	aatgcacaactgatctgacataagaataactctcatctgacaataaataaactaaaccaa	6600
Oy	6601	AAAAAACCC 6608 	
Df	6601	aaaaaaccc 6608	
<hr/>			
RESULT 2			
ID	AA742757	standard: DNA; 153 bp.	
XX	AA742757:		
DT	12-MAR-1997	(first entry)	
DE	Chick CHD-1A gene fragment.		
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1;		
KM	CHD-1A; CHD-W; W chromosome, ss.		
OS	Gallus sp.		
XX	Key	Location/Qualifiers	
FH	misc_difference 52..81	/tag= a	
FT	/note= "bases 52-81 are a repeat of bases 22-51		
FT		and are ignored in the translated amino	
FT		acid sequence given in Fig 3"	
XX	WO9639505-AI.		
XX	12-DEC-1996.		
PD			
PE	05-JUN-1996; 96MO-GB01341.		
PR	06-JUN-1995; 95GB-0011439.		
PA	(ISIS-) ISIS INNOVATION LTD.		
PI	Griffiths R, Tiwari B;		
PB	WPI: 1997-043127/04.		
DR	p-PsDB; AAM08147.		
XX	Avian chromodomain-helicase-DNA binding genes determine sex in		
XX	birds - used for sex determ. and to control sex of progeny		
XX	Claim 8; Fig 3; 76pp; English.		

XX	Bases 3855-3977 (AA742756) of the mouse CHD-1 gene show homology
CC	to portions of the chicken CHD-1A (A = Avian) gene (AA742757).
CC	Chicken CHD-W (W refers to the W chromosome) gene (see also AA742758)
CC	and the great tit CHD-W gene (see also AA742759). Translated amino
CC	acid sequences of this region are provided in AA742754-55. The
CC	CHD-1A (see also AA742751) and CHD-W (see also AA742754-55) genes
CC	determine sex in birds and can be used to identify the sex of an
CC	embryo, locus etc. and to manipulate the sex of progeny.
XX	
SQ	Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other:
OY	
Db	
OY	4100 CAAGAACCCCGGCAAGACGCTCAGACCCTGGCAGACTACCTCATTAATCTAGAA 4159       51 caagaaccccccgcaagcagctacgaccgcgycagactcatctaattactga 110 
OY	4160 TAAAGACCTGCAGAAAGAGACACAAAGCCTTGCTGTGCA 4202 
Db	111 taaagaccttgcaagaagaagacaagaagtgttgcgtgtgca 153 
RESULT	3
AA742754	
ID	AA742754 standard; CDNA; 1316 BP.
XX	
AC	AA742754;
DT	12-MAR-1997 (first entry)
XX	
DE	Chicken CHD-W gene (partial sequence).
XX	
KM	Bird: sex determination; Chromodomain-Helicase-DNA binding 1 Avian;
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX	
OS	Gallus sp.
XX	
PN	WO9639505-A1.
PD	
XX	
PD	12-DEC-1996.
PF	05-JUN-1996; 96MO-GB01341.
PR	06-JUN-1995; 95GB-0011439.
PA	(ISIS-) ISIS INNOVATION LTD.
GR	Griffiths R, Tiwari B;
WP	WPI; 1997-043127/04.
XX	
PT	Avian chromodomain-helicase-DNA binding genes determine sex in
PS	birds - used for sex determ. and to control sex of progeny
Claim 1:	Fig 8; 76pp; English.
XX	
CC	The chicken CHD-W gene (AA742754) acting alone or in conjunction with
CC	the closely related CHD-1A gene (AA742751) is suggested to initiate
CC	female development in birds. The sequence of CHD-W was deduced
CC	from 2 clones isolated from a 10-day chick embryo library using
CC	a fragment of the CHD-1A gene as a probe. The CHD-W gene is
CC	located on the W chromosome. Probes based on CHD-W and CHD-1A give
CC	a W chromosome-specific signal on hybridisation to genomic DNA of a
CC	non-ratite bird and can be used for sex determin. of a bird. CHD-W
CC	nucleic acids can also be used to control the sex of progeny of a
CC	bird.
XX	
SQ	Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other:

Query Match 0.9%; Score 59; DB 18; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 6.3e-16;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2443 ATTATGGGCTAGTCACAAAAGGATCAGTAGAAGAATATCTTTGAAGGCCAAGAA 3001  
|||||  
Db 1 attatcgctagtcacaaaagatcagtagaagaagatattcttgaagaagccaagaa 59

## RESULT 4

AAS85967/C  
ID AAS85967 standard; cDNA; 4222 BP.

AC AAS85967;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #21771.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG21780.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX Claim 1; SEQ ID No 21771; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4222 BP; 1040 A; 1041 C; 883 G; 1258 T; 0 other;

Query Match 0.7%; Score 49; DB 23; Length 4222;  
Best Local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3174 GAGCCCCAGGAAATGATATGATGAATCTTTGAAGAGCTGAAACTC 3222  
|||||  
Db 1765 GAGCCCCAGGAAATGATATGATGAATCTTTGAAGAGCTGAAACTC 1717

## RESULT 5

AAT42758  
ID AAT42758 standard; DNA; 153 BP.

AC AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

XX Key Location/Qualifiers

XX misc\_difference 52..81

XX /tag- a "bases 52-81 are a repeat of bases 22-51  
XX /note- and are ignored in the translated amino  
XX acid sequence given in Fig 3"

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI; 1997-043127/04.

XX P-PSDB; AAW08148.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757) and  
CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
CC sequences of this region are provided in AAW08146-49. The CHD-1A  
CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
CC sex in birds and can be used to identify the sex of an embryo,  
XX foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 0.7%; Score 47; DB 18; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4140 TACCTCATTAATTAATGATAAAGACCTTGCAAGAAAGACGACA 4186  
|||||  
Db 91 tactcatataattactcgataaagaccttgcagaagaagagacaca 137

## RESULT 6

AA142759  
 ID AA142759 standard; DNA; 153 BP.  
 AC AA142759;  
 XX  
 XX  
 DT 12-MAR-1997 (first entry)  
 DE Great tit CHD-W gene fragment.  
 XX  
 XX  
 KM BIRD: sex determination; chromodomain-Helicase-DNA binding 1;  
 KM CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Parus major.  
 XX  
 FH Key Location/Qualifiers  
 FT msc\_difference 52..81  
 FT /tag-<sup>a</sup> bases 52-81 are a repeat of bases 22-51  
 FT /note-<sup>a</sup> and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 FT  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 XX  
 DR P-PSDB; AAM08149.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 XX  
 CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and  
 CC and the great tit CHD-W gene (AA142759).  
 CC sequences of this region are provided in AAM08149. The CHD-1A  
 CC (see also AA142751) and CHD-W (see also AA142754-55) genes determine  
 CC sex in birds and can be used to identify the sex of an embryo,  
 CC foetus etc. and to manipulate the sex of progeny.  
 XX  
 SO Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;  
 XX  
 Query Match 0.64; Score 38; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4140 TACCTGATTAATTAAGTAAGACCTTGCAAGAAA 4177  
 DB 91 tactcattactactgataaagaccttcagaaaa 128  
 XX  
 XX  
 RESULT 7  
 AAC06290  
 ID AAC06290 standard; cDNA; 212 BP.  
 AC AAC06290;  
 XX  
 XX  
 DT 06-OCT-2000 (first entry)  
 DE Human secreted protein 5' EST, SEQ ID NO: 10365.  
 XX  
 XX  
 KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 PN EPI033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 10365; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained for isolating cDNA sequences  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SO Sequence 212 BP; 72 A; 41 C; 32 G; 67 T; 0 other;  
 XX  
 Query Match 0.54; Score 30; DB 21; Length 212;  
 Best Local Similarity 100.0%; Pred. No. 0.0056;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6463 TGGGCAATTAATGCTAATGCTATGCACG 6492  
 DB 5 tgggcaataacatgcataatgctatgcacg 34  
 XX  
 XX  
 RESULT 8  
 ABL32358/c  
 ID ABL32358 standard; DNA; 6237 BP.  
 AC ABL32358;  
 XX  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 331.  
 XX  
 XX  
 KM Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antianemic; cyostatic; noctropic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;  
 KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.

XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
PF  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
DR  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 331; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX  
SQ Sequence 6237 BP; 1614 A; 77 C; 1263 G; 3283 T; 0 other;

Query Match 0.5%; Score 30; DB 24; Length 6237;  
Best Local Similarity 100.0%; Pred. No. 0.0051;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5979 AACACAAACAAACAAACAAACAAATTTAA 6008  
DB 2795 AACACAAACAAACAAACAAACAAATTTAA 2766  
|||||

RESULT 9  
ABL3370/C  
ID ABL3370 standard; DNA: 6191 BP.  
XX  
XX  
AC ABL33370;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX  
DE Human immune system associated gene SEQ ID NO: 1343.

XX Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosstatic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antineumatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX  
OS Homo sapiens.  
XX  
XX WO200200928-A2.  
XX  
XX  
PD 03-JAN-2002.  
XX  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX  
PA (EPIC-) EPIGENOMICS AG.  
XX

PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
DR  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 1343; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX  
SQ Sequence 6191 BP; 1678 A; 92 C; 1597 G; 2824 T; 0 other;

Query Match 0.4%; Score 29; DB 24; Length 6191;  
Best Local Similarity 100.0%; Pred. No. 0.014;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5979 AACACAAACAAACAAACAAACAAATTTAA 6007  
DB 328 AACACAAACAAACAAACAAACAAATTTAA 300  
|||||

RESULT 10  
ABL33795/C  
ID ABL33795 standard; DNA: 5474 BP.  
XX  
XX  
AC ABL33795;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX  
DE Human immune system associated gene SEQ ID NO: 1768.

XX Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosstatic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX  
OS Homo sapiens.  
XX  
XX WO200200928-A2.  
XX  
XX  
PD 03-JAN-2002.  
XX  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX  
PA (EPIC-) EPIGENOMICS AG.  
XX  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
DR  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 1768; 32pp + Sequence Listing; German.  
XX

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/alcerative bowel diseases. The present sequence is a gene of the invention.

Sequence 5474 BP; 1286 A; 201 C; 1404 G; 2583 T; 0 other;

Query Match                      0.4%; Score 27; DB 24; Length 5474;  
Best Local Similarity    100.0%; Fred. No. 0.11;  
Matches    27; Conservative    0; Mismatches    0; Indels    0; Gaps    0

Oy    5979 AACCAACAAACAACAAAAAACATT    6005  
     | | | | | | | | | | | | | | | | |  
Db    2887 AACCAACAAACAACAAAAAACATT    2861

RESULT    11  
AAZ13346  
ID    AAZ13346 standard; cDNA: 300 BP.  
XX  
AC    AAZ13346;  
XX  
DT    12-OCT-1999 (first entry)  
XX  
DE    Human gene expression product cDNA sequence SEQ ID NO:815.  
XX  
KW    Human; gene: gene expression product; diagnosis; therapy; probe:  
KM    detection; mapping; tissue typing; profiling; forensic; cancer:  
XX    genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX  
OS    Homo sapiens.  
CS  
PN    WO938972-A2.  
PD  
PD    05-AUG-1999.  
PR  
PR    28-JAN-1999;    99WO-US01619.  
XX  
FR    03-APR-1998;    98US-0080666.  
XX  
FR    28-JAN-1998;    98US-0072910.  
PR    24-FEB-1998;    98US-0075954.  
PR    31-MAR-1998;    98US-0080114.  
PR    03-APR-1998;    98US-0080515.  
XX  
PA    (CHIR) CHIRON CORP.  
XX    (HYSF-) HYSEQ INC.  
PI  
PI    Escobedo J, Garcia M, Drmanac R, Drmanac S;  
CC    Escobedo J, Garcia PD, Garcia V, Glase K, Innis MA;  
CC    Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
PI    Lomson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI    Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX  
DR    WPI: 1999-494092/41.  
PT  
PT    Novel human genes and their expression products which are  
XX    differentially expressed in different cell types  
PS  
PS    Claim 1: Page 849; 2479pp; English.

The present invention describes a library of human polynucleotides comprising the sequences given in AA212532 to AA21779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA212532 to AA21779. The polynucleotides can be used as a source of primers and probes, which can

	CC	be used for a variety of purposes-e.g. detection of expression levels,
	CC	mapping, tissue typing or profiling, forensics, genetic analysis and
	CC	detection of polymorphisms. Polypeptides encoded by the polynucleotides
	CC	can be used for raising antibodies for experimental, diagnostic and
	CC	therapeutic purposes. The polynucleotides may also be used to construct
	CC	arrays for diagnostics (which may be used to determine function of an
	CC	encoded protein) and to detect differences in expression levels between
	CC	two cells (e.g. to identify abnormal or diseased tissue in a human, to
	CC	identify a genetic predisposition or susceptibility to a disease such as
	CC	cancer). The polynucleotides of the invention are especially used in the
	CC	diagnosis, prognosis and management of colorectal cancer, breast cancer,
	CC	and lung cancer. The polynucleotides can also be used to screen for
	CC	peptide analogues and antagonists.
SQ		
Sequence	300 BP; 113 A; 66 C; 59 G; 62 T; 0 other;	
Query Match	0.4%; Score 26; DB 20; Length 300;	
Best Local Similarity	100.0%; Pred. No. 0.34;	
Matches	26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	5979 AACCAAAACAACAAAACAAAAAAGAT 6004	
Db	30 aaacaacaaacacaaaatacat 55	
RESULT 12		
AAK75594		
ID	AAK75594 standard; DNA; 2785 BP.	
AC	AAK75594;	
XX		
DT	07-NOV-2001 (first entry)	
XX		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30406.	
XX		
KW	Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KM	Cytostatic; gene therapy; vaccine; metastasis; ds.	
OS	Homo sapiens.	
FN	WO200157182-A2.	
PD	09-AUG-2001.	
FE	17-JAN-2001; 2001WO-USO1354.	
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216680.	
PR	07-JUL-2000; 2000US-0216887.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217486.	
PR	14-YUL-2000; 2000US-0218920.	
PR	26-YUL-2000; 2000US-0220963.	
PR	26-YUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224318.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	

PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0234984.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 01-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249287.  
 PR 17-NOV-2000; 2000US-0249289.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-483426/52.  
 DR  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 XX  
 PS Disclosure; SEQ ID NO 30406; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 2785 BP; 840 A; 572 C; 594 G; 779 T; 0 other;

Query Match 0.48; Score 26; DB 22; Length 2785;  
 Best Local Similarity 100.0%; Pred. No. 0.32;



Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5977 TGAACAACAACAACAACAACA 6002  
|||||  
Db 192 TGAACAACAACAACAACAACA 217

## RESULT 13

AA546349/C  
ID AA546349 standard; DNA; 5919 BP.

AC AA546349;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #71.

KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

KW cytosine methylation; ds.

OS Homo sapiens.

PN WO200168912-A2.

PD 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI: 2001-602752/68.

PT Fragments of chemically modified genes associated with tumour suppressor

PT genes and oncogenes; useful in designing primers and probes for

PT analysing diseases associated with cytosine methylation state e.g.

PT cancer

PS Claim 1; SEQ ID NO 71; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18

CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and

CC oncogenes having a sequence taken from 536 (actually 533 since

CC (SS) and sequences complementary to (SS). The nucleic acid may be a

CC form part of a set of probes for detecting the cytosine methylation state

CC and/or single nucleotide polymorphisms and also to be used in an

CC array for analysing diseases associated with CpG dinucleotides e.g.

CC cancers and tumours. The probes can also be used in a method for

CC ascertaining genetic and/or epigenetic parameters for the diagnosis

CC and/or therapy of existing diseases or the predisposition to specific

CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the

CC are disadvantageous to patients. The present sequence is one of the

CC 533 genomic sequences derived from tumour suppressor genes and

CC oncogenes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 5919 BP; 1421 A; 69 C; 1641 G; 2788 T; 0 other;

Query Match 0.4%; Score 26; DB 22; Length 5919;

Best Local Similarity 100.0%; Pred. No. 0.31;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5979 AAACAACAACAACAACAACA 6004

|||||

Db 3118 AAACAACAACAACAACAACA 3093

## RESULT 14

ABL32732/C  
ID ABL32732 standard; DNA; 5919 BP.

AC ABL32732;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 705.

KW Human; immune system disease; cytosine methylation; antitumour;

KW antiarteriosclerotic; antianemic; cytostatic; neoplastic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antidiabetic; antipsoriatic;

KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 705; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 5919 BP; 1421 A; 69 C; 1641 G; 2788 T; 0 other;

XX Query Match 0.4%; Score 26; DB 24; Length 5919;

XX Best Local Similarity 100.0%; Pred. No. 0.31;

XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 5979 AAACAACAACAACAACAACA 6004

XX |||||||

XX Db 3118 AAACAACAACAACAACAACA 3093

RESULT 15  
AAS61168/c  
ID AAS61168 standard; DNA: 5919 BP.  
AC AAS61168;

DT 29-JAN-2002 (first entry)

DE Human gene regulation-associated gene oligonucleotide #123.

XX Human: Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preeclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

OS Homo sapiens.

PN WO200177375-A2.

XX 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP03968.

XX 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Plegenbrock C, Berlin K;

DR WPI; 2002-017470/02.

XX New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease -  
XX Claim 1; SEQ ID No 127; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preeclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5919 BP; 1421 A; 69 C; 1641 G; 2788 T; 0 other;

OY 5979 AACCAACCAACCAACCAACCAAT 6004  
|||  
Db 3118 AACCAACCAACCAACCAACCAAT 3093


Search completed: August 3, 2002, 01:52:38  
Job time: 32343 sec

Query Match 0.4%; Score 26; DB 24; Length 5919;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Mon Aug 5 11:51:35 2002

us-08-973-363-10.oli.rng

Page 13



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:40 ; Search time 7016.61 Seconds

(without alignments)  
12710.964 Million cell updates/sec

Title: US-08-973-363-10

Perfect score: 6608

Sequence: 1 CGCGCTCGCGACGAGAGCGC.....AACTAAACCAAAAAAACCC 6608

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched:

Word size : 0  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlin:\*  
5: em\_estlin:\*  
6: em\_estlin:\*  
7: em\_estlin:\*  
8: em\_estlin:\*  
9: gb\_estl:\*  
10: gb\_estl:\*  
11: gb\_estl:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	443	6.7	597	10	BM486590
2	394	6.0	648	10	BM486590
3	87	1.3	476	10	BE479788
4	82	1.2	234	9	AM075773
5	82	1.2	252	9	AA824571
6	82	1.2	296	9	AA80460
7	82	1.2	354	9	AA83709
8	82	1.2	399	10	H99726
9	82	1.2	400	10	H99726
10	82	1.2	408	9	AM591975
11	82	1.2	419	9	AA42118
12	82	1.2	434	9	AA609977
13	82	1.2	436	9	AA609977
14	82	1.2	453	9	AA609977
15	82	1.2	466	9	BE048526
16	82	1.2	490	9	AA599181
17	82	1.2	490	9	AA599181

18	82	1.2	505	9	AI298448
19	82	1.2	642	10	BE873688
20	82	1.2	769	10	BE813057
21	82	1.2	802	10	BM015816
22	82	1.2	1280	10	BM015816
23	79	1.2	945	10	BE535223
24	74	1.1	422	10	R44604
25	67	1.0	543	9	BE559370
26	67	1.0	543	9	BE559370
27	67	1.0	551	9	AI109754
28	67	1.0	566	9	AI176714
29	67	1.0	597	9	AI106778
30	67	1.0	613	9	AA943664
31	67	1.0	637	9	BE107423
32	66	1.0	637	9	AU045087
33	66	1.0	353	9	AU045087
34	66	1.0	353	9	AU045087
35	66	1.0	442	9	AA273291
36	66	1.0	486	9	AA116690
37	66	1.0	495	10	BE447654
38	66	1.0	531	10	BE194545
39	66	1.0	1346	11	AK018451
40	63	1.0	238	10	BE704506
41	63	1.0	343	9	AA885644
42	63	1.0	401	10	BE550103
43	63	1.0	509	10	BE813029
44	63	1.0	539	10	BE812861
45	63	1.0	549	10	BE813014

## ALIGNMENTS

RESULT 1  
LOCUS BM486590  
DEFINITION pgm2n.pk002.b16 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA clone pgm2n.pk002.b16 5' similar to gblAAC60282.1 (AF004337)  
chromo-helicase-DNA-binding on the z chromosome protein [Gallus gallus], mRNA sequence.  
ACCESSION BM486590  
VERSION BM486590.1 GI:18607520  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 597)  
Cognburn, L.A. and Monson-O'Connell, E.  
ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library, USDA/IFARS Animal Genome Project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Larry A. Cognburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cognburn@udel.edu, www.chickest.udel.edu.

## FEATURES

source  
1..597  
/organism="Gallus gallus"  
/strain="Commercial broiler and Ottawa Res. Centre Strains 90 & 21"  
/db\_xref="taxon:9031"  
/clone\_lib="Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n)"  
/sex="Male and Female"  
/tissue\_type="Breast muscle, Leg muscle and epiphyseal growth plate"

Query Match	Best Local Similarity	Score	DB	Length
Matches 563; Conservative 0; Mismatches 0; Indels 1; Gaps 1;	99.8%;	Pred. No. 1,1e-173;		597;
QY 5172	TCGACCTGAGATATACAGCCATCATATAATCTTCGAGAGATTATAGATATACACACTGAGACTGG	5231		
Db 1	TCGACCTGAGATATACAGCCATCATATAATCTTCGAGAGATTATAGATATACACACTGAGACTGG	60		
QY 5232	CAATGAGACACAGAGCTTCTGGTAGTGCCCGAGGTACACACTAGATCAGAGCTCTCT	5291		
Db 61	CAATGAGACACAGAGCTTCTGGTAGTGCCCGAGGTACACACTAGATCAGAGCTCTCT	120		
QY 5292	TATGGTTCAAGATCTCCCTAGGACACAGATCTCCATTGTAACACTCATCATGATACAAA	5351		
Db 121	TATGGTTCAAGATCTCCCTAGGACACAGATCTCCATTGTAACACTCATCATGATACAAA	180		
QY 5352	AGTACACCTGAAACATATACATGAGTGGCCGGAAGACATATAACAAAGACTGACATTTTCTGGA	5411		
Db 181	AGTACACCTGAAACATATACATGAGTGGCCGGAAGACATATAACAAAGACTGACATTTTCTGGA	240		
QY 5412	CCTTTCTTTTAGCCATATACAGTAAACTAACAGTAATGGCTTTACATGACTTGAAGA	5471		
Db 241	CCTTTCTTTTAGCCATATACAGTAAACTAACAGTAATGGCTTTACATGACTTGAAGA	300		
QY 5472	TATGAGACGTGATATCTATCATAGTACAGATATTTGTTACTCTTCCAGATGCAAGGTCTA	5531		
Db 301	TATGAGACGTGATATCTATCATAGTACAGATATTTGTTACTCTTCCAGATGCAAGGTCTA	360		
QY 5532	TTATCCCAACAGAAAGAAATATTTTGTATTTAAAGTTATATGTCGACATGCTGCACAA	5591		
Db 361	TTATCCCAACAGAAAGAAATATTTTGTATTTAAAGTTATATGTCGACATGCTGCACAA	420		
QY 5592	TGTTGTGCGACTTTTATTTTAAAGAAATGGAAGATGTTTACTTTTACAGGAGCCTCAACAC	5651		
Db 421	TGTTGTGCGACAC-TTTTATTTTAAAGAAATGGAAGATGTTTACTTTTACAGGAGCCTCAACAC	479		
QY 5652	TGCCCCCTTCAGACGTGAGATCTTACTATATAAATCTTCAATGTCAAAGTGTTCTAGCTGGA	5711		
Db 480	TGCCCCCTTCAGACGTGAGATCTTACTATATAAATCTTCAATGTCAAAGTGTTCTAGCTGGA	539		
QY 5712	ACACAGATTAAATTTATGTTTGTAA	5735		
Db 540	ACACAGATTAAATTTATGTTTGTAA	563		
RESULT 2				
LOCUS	BM491730	648 bp	mRNA	linear
DEFINITION	pBp2n.pk007.e18 Normalized Chicken Pituitary/Hypothalamus/Pineal Library (pBp2n.gallus gallus cdna clone pBp2n.pk007.e18 5' similar to gblfAA6502822.1 (AF004397) chromo-helicase-DNA-binding on the 2 chromosome protein [Gallus gallus], mRNA sequence.			
ACCESSION	BM491730			
VERSION	BM491730.1	GI:18612661		
KEYWORDS	EST.			
SOURCE	chicken.			
ORGANISM	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
	Phasianinae; Gallus			
REFERENCE	1 (bases 1 to 648)			

AUTHORS	Porter,T.E. and Coopburn,L.A.
TITLE	ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal CDNA Library, USDA/NIH's Animal Genome Project
JOURNAL	Unpublished (2002)
COMMENT	Contact: Larry A. Coopburn University of Delaware Townsend Hall, Newark, DE 19717, USA Tel: 302-831-1335 Fax: 302-831-2822 Email: coopburnudel.edu, www.chickest.udel.edu.
FEATURES	Location/Qualifiers
source	1..648
	/organism="Gallus gallus"
	/strain="Commercial broiler chickens"
	/db_xref="taxon:9031"
	/clone_pgp2n.pk007.e18"
	/clone_lib="Normalized Chicken
	Pituitary/Hypothalamus/Pineal Library (pgp2n)"
	/sex="Male and Female"
	/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
	/dev_stage="Embryonic (dl2,d14,d19) : post-hatch (1,3,5,7,5 weeks)"
	/lab_host="E. coli EMDH10B"
	/note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"
BASE COUNT	203 a 116 c 127 g 190 t 12 others
ORIGIN	
Query Match	6.0% ; Score 394; DB 10; Length 648;
Best Local Similarity	99.5%; Pred. No. 2.9e-153;
Matches 544; Conservative	0; Mismatches 3; Indels 0; Gaps 0.
OY	1493 TGGTCTCTCATATGCCAAAAAGTTTCAGGACGCATTGTAGATATTTTAGCAGAATCA 1552
Db	1 TGGTGCTCTCATATGCCAAAAAGTTTCAGGACGCATTGTAGATATTTTAGCAGAATCA 60
OY	1553 ATCCAAAGACTACTCCCTTTAAGGACTGCAGAGTTCTAAAACAGAACCAAGATTTGGTGC 1612
Db	61 ATCCAAAGACTACTCCCTTTAAGGACTGCAGAGTTCTAAAACAGAACCAAGATTTGGTGC 120
OY	1613 ACTGAAGAAGCACCATCTTACATTGGAGACATGAAGTCTGGAGTTTAAGAGATTATCA 1672
Db	121 ACTGAAGAAGCACCATCTTACATTGGAGACATGAAGTCTGGAGTTTAAGAGATTATCA 180
OY	1673 GTTAATAGGATGAAWTGGCTCGCTCATTCATGATGTCAGAAAGAAATAGTTATCTTGC 1732
Db	181 GTTAATAGGATGAAWTGGCTCGCTCATTCATGATGTCAGAAAGAAATAGTTATCTTGC 240
OY	1733 AGATGAATGGGTCTGGGTAAACAATACAAACAATTTCTTTCTGMACTACTCGTTTCA 1792
Db	241 AGATGAATGGGTCTGGGTAAACAATACAAACAATTTCTTTCTGMACTACTCGTTTCA 300
OY	1793 TGAACATCACTGTATGGCCCTTTTCTTCTTGGCGGTCGCACATTTCCTTACATCTTG 1852
Db	301 TGAACATCACTGTATGGCCCTTTTCTTCTTGGCGGTCGCACATTTCCTTACATCTTG 360
OY	1853 GCAAAGAGAGATTCAAACTTGGGCTCCAGATGAAATGCTGTAGTTTACTTAGGAGATAT 1912
Db	361 GCAAAGAGAGATTCAAACTTGGGCTCCAGATGAAATGCTGTAGTTTACTTAGGAGATAT 420
OY	1913 AACCTAGTGAATATGATTAAGACTCATGAATGATGCATCCACAGACTAAACGATTAAA 1972
Db	421 AACCTAGTGAATATGATTAAGACTCATGAATGATGCATCCACAGACTAAACGATTAAA 480
OY	1973 GTTTAACATACTTTCAGACACATATGAATTTTACTGAAGGATTAAGTCAATTCCTGGTGG 2032
Db	481 GTTTAACATACTTTCAGACACATATGAATTTTACTGAAGGATTAAGTCAATTCCTGGTGG 540
OY	2033 TCTCAAT 2039
Db	541 TCTCAAT 547

RESULT 3  
LOCUS BE479788 476 bp mRNA linear EST 28-AUG-2000  
DEFINITION 164616 BARC 580V Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE479788  
VERSION BE479788.1 GI:95959321  
KEYWORDS EST.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 476)  
AUTHORS Sonstegard, T.S., Capuco, A.V., Van Tassell, C.P., Ashwell, M.S. and Wells, K.D.  
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8414  
Email: tads@psl.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTACGACG  
Plate: 14 row: E column: 24  
Seq primer: ATTTAGTGACACTATAG.  
FEATURES  
source  
1..476  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="BARC 580V"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."  
BASE COUNT 137 a 94 c 67 g 178 t  
ORIGIN  
Query Match 1.3%; Score 87; DB 10; Length 476;  
Best Local Similarity 100.0%; Pred. No. 3.3e-25;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 5831 TTCACTTTCCAAAGCTACTGTTTACATGTGACACGACACCTTCCGCTTTTCATC 5890  
DB 183 TTCACTTTCCAAAGCTACTGTTTACATGTGACACGACACCTTCCGCTTTTCATC 242  
OY 5891 ACACGCTGAATATTTAAATCTGTAC 5917  
DB 243 ACACGCTGAATATTTAAATCTGTAC 269  
RESULT 4  
LOCUS AM075773 234 bp mRNA linear EST 13-OCT-1999  
DEFINITION xabsc01.x1 NCI\_CGAP\_CML1 Homo sapiens cDNA clone IMAGE:2573568 3',  
mRNA sequence.  
ACCESSION AM075773  
VERSION AM075773.1 GI:6030771  
KEYWORDS EST.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 234)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,  
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life  
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/btrp/image/image.html](http://www.bio.lnl.gov/btrp/image/image.html)  
Seq primer: -400P from Gibco.  
FEATURES  
source  
1..234  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2573568"  
/clone\_lib="NCI\_CGAP\_CML1"  
/tissue\_type="myeloid cells, 18 pooled CML cases, BCR/ABL  
rearrangement positive, includes both chronic phase and  
myeloid blast crisis"  
/lab\_host="DH10B"  
/note="Organ: whole blood; Vector: PCMV-SPORT6; Site.1:  
SalI; Site.2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Library constructed by Life Technologies."  
BASE COUNT 88 a 28 c 46 g 72 t  
ORIGIN  
Query Match 1.2%; Score 82; DB 9; Length 234;  
Best Local Similarity 100.0%; Pred. No. 5.5e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 5836 TTTCNCAAGCTACTGTTTACATGTGACACGACACCTTCCGCTTTTCATCAG 5895  
DB 198 TTTCNCAAGCTACTGTTTACATGTGACACGACACCTTCCGCTTTTCATCAG 139  
OY 5896 CTGGATTTTAAATCTGTAC 5917  
DB 138 CTGGATTTTAAATCTGTAC 117  
RESULT 5  
LOCUS AA824571 252 bp mRNA linear EST 07-APR-1998  
DEFINITION oc78h11.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1355877,  
mRNA sequence.  
ACCESSION AA824571  
VERSION AA824571.1 GI:2896593  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 252)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Louis M. Straudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at  
www-bio.lnl.gov/brp/image/image.html  
Insert Length: 1636 Std Error: 0.00  
Seq primer: -40m13 fwd. EF from Amersham  
High quality sequence stop: 83.

**FEATURES**  
**source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1355877"
/clone_lib="NCI-CGAP-GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
cDNA was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Straut (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - Oligo(dT) primer
[5'-TGTTACCAATCTGACAGCGAGCGCGCCCTCATTTTTTTTTTTTTT-3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73D vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

BASE COUNT  
ORIGIN

85 a	29 c	41 g	97 t
------	------	------	------

Query Match	1.2%	Score 82;	DB 9;	Length 252;
Best Local Similarity	100.0%	Pred. No. 5.3e-23;		
Matches 82; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	5836	CTTGAAATATTAATCTGTAC	5917
QY	5836	TTTCCAAAGCTACTGTTTACATGTGACACCTGGACACTTGCCTTTTCATCACAG	5895
Db	211	TTTCCAAAGCTACTGTTTACATGTGACACCTGGACACTTGCCTTTTCATCACAG	152
QY	5896	CTTGAAATATTAATCTGTAC	5917
Db	151	CTTGAAATATTAATCTGTAC	130

RESULT	6
LOCUS	AA480460
DEFINITION	AA480460 296 bp mRNA linear EST 14-AUG-1997 ne010d08.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:903615, mRNA sequence.
ACCSSION	AA480460
VERSION	AA480460.1 GI:2208611
KEYWORDS	EST.
SOURCE	human. tissue: human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 296)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D.

cDNA Library Preparation: David B. Kitzman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E Consortium/LLNL at:  
[www.bio.llnl.gov/bbnp/image/image.html](http://www.bio.llnl.gov/bbnp/image/image.html)  
 Insert Length: 280 Std Error: 0.00  
 Seq primer: -4im13 fwd. ET from Amersham  
 High quality sequence stop: 288.

**FEATURES**  
**SOURCE**

```

FEATURES
    source
        location/Qualifiers
            1..296
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:909615"
                /clone_1ib="NCI-CGAP_Ew1"
                /tissue_type="Ewing's sarcoma"
                /lab_host="DH10B"
                /note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
                cDNA made by oligo-dT priming. Non-directionally cloned.
                Size-selected on agarose gel, average insert size 600 bp
                Reference: Krizman et al. (1996) Cancer Research
                56:5380-5383."
BASE COUNT
    87 a      54 c      45 g      110 t
ORIGIN

```

BASE COUNT	87 a	54 c	45 g	110 f
ORIGIN				

87 a	54 c	45 g	110 t
------	------	------	-------

Query Match	1.2%;	Score 82;	DB 9;	Length 296;
Best Local Similarity	100.0%;	Pred. No. 4.9e-23;		
Matches 82;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	5836	TTTCCAAGC	ACTGTTGTTTACATGTCACACGCACACACCCTGGCGGTTTTTCATCACAAG	5887
QY	5836	TTTCCAAGC <td>ACTGTTGTTTACATGTCACACGCACACACCCTGGCGGTTTTTCATCACAAG</td> <td>5887</td>	ACTGTTGTTTACATGTCACACGCACACACCCTGGCGGTTTTTCATCACAAG	5887
Db	158	TTTCCAAGC <td>ACTGTTGTTTACATGTCACACGCACACACCCTGGCGGTTTTTCATCACAAG</td> <td>217</td>	ACTGTTGTTTACATGTCACACGCACACACCCTGGCGGTTTTTCATCACAAG	217
QY	5896	CTTGAAATATTTAAATCTGTAC	5917	
Db	218	CTTGAAATATTTAAATCTGTAC	239	

	TITLE	H80753	352 bp	mRNA	linear	EST_09-NOV-1996
	JOURNAL	yv78d12.s1 Soares melanocyte 2Nbhm Homo sapiens cDNA clone IMAGE:24885 3', mRNA sequence.				
	COMMENT	H80753				
	REFERENCE	H80753				
	AUTHORS	GI:1058842				
	SOURCE	human.				
	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 352) Hillier,L., Clark,N., Dubucque,T., Elliston,K., Hawkins,M., Holman, M., Hultman,M., Kneaba,T., Le M., Lennon,G., Marra,M., Parsons,J., Rikkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston , R., Williamson,A., Woldmann,P. and Wilson,R. The Washu-Merck EST Project Unpublished (1995) Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 224 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium ( <a href="mailto:infoimage.lnl.gov">infoimage.lnl.gov</a> ) for further information. Seq primer: Promega -21ml3 High quality sequence stop: 224.				

**FEATURES**

**SOURCE**

```

/organism="Homo sapiens"
/db_xref="GDB:3866561"
/db_xref="taxon:9606"
/clone_image="248855"
/clone_11b="Soares melanocyte 2NBHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10 (ampicillin resistant)"
/notes="Vector: pU73D (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA"
```



was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTGACAGTGGAGCGGCGGCTTTTTCATCAGAG 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pRT3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
 (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 123 a 56 c 68 g 102 t 3 others

Query Match 1.2% Score 82; DB 10; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 5895  
 DB 139 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 80  
 OY 5896 CTTGAATATTTAAATCTGTAC 5917  
 DB 79 CTTGAATATTTAAATCTGTAC 58

RESULT 8  
 LOCUS AA483709 354 bp mRNA linear EST 15-AUG-1997  
 DEFINITION ne75g04.s1 NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:910134, mRNA  
 accession  
 AA483709  
 VERSION AA483709.1 GI:2212522  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 354)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.

CDNA Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/bprp/image/image.html  
 Seq primer: -41m3 fwd. RT from Amersham.

FEATURES  
 source

1.354  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:910134"  
 /clone\_1lb="NCI\_CGAP\_Ew1"  
 /tissue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP10; mRNA made from Ewing's sarcoma,  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size-selected on agarose gel, average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 124 a 58 c 53 g 119 t  
 ORIGIN  
 Query Match 1.2% Score 82; DB 9; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 5895  
 DB 140 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 199  
 OY 5896 CTTGAATATTTAAATCTGTAC 5917  
 DB 200 CTTGAATATTTAAATCTGTAC 221

RESULT 9  
 LOCUS H99736 399 bp mRNA linear EST 15-DEC-1995  
 DEFINITION yx27g03.s1 Soares melanocyte 2NBHM Homo sapiens cDNA clone  
 IMAGE:262996 3', mRNA sequence.  
 ACCESSION H99736  
 VERSION H99736.1 GI:1124404  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 399)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston,  
 R., Williamson, A., Woldman, P., and Wilson, R.  
 The Wash-Merck EST Project  
 Unpublished (1995)  
 CONTACT: Wilson R.  
 JOURNAL  
 COMMENT Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

Source: IMAGE Consortium, LNL  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 968 Std Error: 0.00  
 Seq primer: m13 -40 forward  
 High quality sequence stop: 400.

FEATURES  
 source

1.399  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:262996"  
 /clone\_1lb="Soares melanocyte 2NBHM"  
 /sex="Male"  
 /tissue\_type="melanocyte"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pRT3D (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTGACAGTGGAGCGGCGGCTTTTTCATCAGAG 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pRT3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
 (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 140 a 60 c 73 g 126 t  
 ORIGIN  
 Query Match 1.2% Score 82; DB 10; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 5895  
 DB 191 TTTCGAAGCTACTGTTTACATGTCAGTGGAGCACCCTTGGCGTTTCATCAGAG 132

QY 5896 CTTGAATATTTAAATTCGTAC 5917  
|||||  
LOCUS 131 CTTGAATATTTAAATTCGTAC 110

RESULT 10  
H99724/c 400 bp mRNA linear EST 15-DEC-1995  
LOCUS  
DEFINITION YX27e03.s1 Soares melanocyte 2NBHM Homo sapiens cDNA clone  
IMAGE:262972.3', mRNA sequence.

ACCESSION H99724  
VERSION H99724.1 GI:1124392  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 400)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 360  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 929 Std Error: 0.00  
Seq primer: m13 -40 forward  
High quality sequence stop: 360.  
Location/Qualifiers  
1. 400  
/organism="Homo sapiens"  
/db\_xref="GDB:3872614"  
/db\_xref="taxon:9606"  
/clone="IMAGE:262972"  
/clone\_1lb="Soares melanocyte 2NBHM"  
/sex="Male"  
/tissue\_type="melanocyte"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT773D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCGCGAGTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT773 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 139 a 61 c 73 g 126 t 1 others  
ORIGIN

Query Match 1.28; Score 82; DB 10; Length 400;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCACAAAGCTACTGTTTACATGTACACGACACCTTGGCGTTTCATCACACAG 5895  
|||||  
Db 191 TTTCACAAAGCTACTGTTTACATGTGTACACGACACCTTGGCGTTTCATCACACAG 132

QY 5896 CTTGAATATTTAAATTCGTAC 5917  
|||||  
Db 131 CTTGAATATTTAAATTCGTAC 110

RESULT 11  
AM591975/c 408 bp mRNA linear EST 22-MAR-2000  
LOCUS  
DEFINITION h36e06.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2933986.3', mRNA sequence.

ACCESSION AM591975  
VERSION AM591975.1 GI:7279145  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 408)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: ccgaps@mail.nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -400P from GIDCO  
High quality sequence stop: 356.  
Location/Qualifiers  
1. 408  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2933986"  
/clone\_1lb="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT773D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NRT, and B-cell NCI-CCAP-CCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 150 a 53 c 77 g 128 t  
ORIGIN

Query Match 1.28; Score 82; DB 9; Length 408;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCACAAAGCTACTGTTTACATGTACACGACACCTTGGCGTTTCATCACACAG 5895  
|||||  
Db 273 TTTCACAAAGCTACTGTTTACATGTGTACACGACACCTTGGCGTTTCATCACACAG 214

QY 5896 CTTGAATATTTAAATTCGTAC 5917  
|||||  
Db 213 CTTGAATATTTAAATTCGTAC 192

RESULT 12  
AA442118 419 bp mRNA linear EST 02-JUN-1997  
LOCUS  
DEFINITION zw55h08.r1 Soares total fetus NB2HFB\_9w Homo sapiens cDNA clone  
IMAGE:774015.5', mRNA sequence.

ACCESSION AA442118  
VERSION AA442118.1 GI:2153996  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 419)



FEATURES  
source  
Location/Qualifiers  
1.453  
/organism="Homo sapiens"  
/db\_xref="GDB:3873587"  
/db\_xref="taxon:9606"  
/clone="IMAGE:263945"  
/sex="Male"  
/tissue.type="melanocyte"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(CT) primer [5',  
TCTTACCAATCTGAGTGGGCGCGGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 163 a 68 c 83 g 139 t

ORIGIN

Query Match 1.2%; Score 82; DB 10; Length 453;  
Best Local Similarity 100.0%; Pred. No. 4.1e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCGAAGCTACTGTTTACATGTACACGCGACCCCTTGGCGCTTTTCATCACAG 5895  
|||||  
Db 191 TTTCGAAGCTACTGTTTACATGTACACGCGACCCCTTGGCGCTTTTCATCACAG 132  
|||||

QY 5896 CTGGAATATTTAAATCTGTAC 5917  
|||||  
Db 131 CTGGAATATTTAAATCTGTAC 110  
|||||

RESULT 15  
LOCUS A1088996 456 bp mRNA linear EST 10-NOV-1998  
DEFINITION IMAGE:1634841 3', mRNA sequence.  
ACCESSION A1088996  
VERSION A1088996.1 GI:3428055  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 456)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LINT; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 1133 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 449.  
Location/Qualifiers  
1.456  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1634841"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in

FEATURES  
source

a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
309384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HP-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326653 Soares NBHOR  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 166 a 63 c 92 g 135 t

ORIGIN

Query Match 1.2%; Score 82; DB 9; Length 456;  
Best Local Similarity 100.0%; Pred. No. 4.1e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCGAAGCTACTGTTTACATGTACACGCGACCCCTTGGCGCTTTTCATCACAG 5895  
|||||  
Db 257 TTTCGAAGCTACTGTTTACATGTACACGCGACCCCTTGGCGCTTTTCATCACAG 198  
|||||

QY 5896 CTGGAATATTTAAATCTGTAC 5917  
|||||  
Db 197 CTGGAATATTTAAATCTGTAC 176  
|||||


Search completed: August 2, 2002, 22:41:57  
Job time: 30172 sec

Mon Aug 5 11:51:36 2002

us-08-973-363-10.oli.rst

---

Page 9



**THIS PAGE BLANK (USPTO)**



NKEKREKVKESTQKEKEVKEVNEKSKSKKIPLLDTPVHTTATSEPYPIS  
ESELHOKITFSVCKERMYPKAAKQIDRPEKGLSEBEOLEHTPOCLIKIGDHTTECL  
KEYTNPEOIKOMRKNLWTFYSKFTFEPDARLHLKLYAIKKROSOOHONONISSNVN  
THYIRNDPDRLEKETTNDHSDSDSYSDSHLSQYHDHMDRDRHOGDAYKKSDDSKRRY  
SATSNGDHDMDHKODSKRYSDSKSRKIRLDDHRSRHRSLBENLKDSDGSHSRSH  
SDRHSDDHRSSTSYSHKSSRDYRSHDQMDHRASGSPRSLDQSPYSPLG  
HRSPEHSSDHKSTPEHTWSSRKT\*  
misc\_feature  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

## alignment\_scores:

Quality: 88.00 Length: 88  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 AspGluIleValSerValIshIseuHisLysLysIleLysThrGluLys 17  
|||||  
4341 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATTAACAGAAAA 4390  
17 sgluysngluylusprogluproaspilleglytlelysylsgluylag 34  
|||||  
4391 AGAAATGAGAAAAAGCCTGAGCCAGATATTGTTAAAGAGAAAGCTG 4440  
34 luGluLysArgGluThrLysGluLysGluLysArgGluLysArg 50  
|||||  
4441 AAGAAAAAGAGACAGACAAAAAGAAAGAAATTAAGGCAATTGAAAAAG 4490  
51 GluLysLysGluLysGluLysLysLysGluLysGluLysAspAsnLys 67  
|||||  
4491 GGAAAAAGAAAAAGAGCATTAAGAAATTAAGAAAAAGAAATTAATA 4540  
67 sgluysarggluysnlysvallysgluSerThrGlnLysGluLysGlu 84  
|||||  
4541 AGAAAAAGAGAAAAACAAAGTAAGAAATCCACACAGAAAGAAAAAGAG 4590  
84 alylsgluGluLys 88  
|||||  
4591 TGAAGGAGAGAGAG 4604

seq\_name: gb\_pat:A58693

## seq\_documentation\_block:

LOCUS A58693 265 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 12 from Patent WO9639505.  
ACCESSION A58693  
VERSION A58693.1 GI:3714251  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE 1 (bases 1 to 265)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 12 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES  
Location/Qualifiers  
source  
1..265  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 158 a 16 c 61 g 30 t  
ORIGIN

## alignment\_scores:

Quality: 75.00 Length: 89  
Ratio: 0.852 Gaps: 1  
Percent Similarity: 98.876 Percent Identity: 98.876

## alignment\_block:

US-08-973-363-11 x A58693 ..

Align seg 1/1 to: A58693 from: 1 to: 265

1 AspGluIleValSerValIshIseuHisLysLysIleLysThrGluL 17  
|||||  
1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATTAACAGAAAA 50  
17 ysgluysngluylusprogluproaspilleglytlelysylsgluila 33  
|||||  
51 AAGAAAAAGAGAAAAAGCCTGAGCCAGATATTGTTAAAGAGAAAGCT 100  
34 GluGluLysArgGluThrLysGluLysGluLysArgGluLysArg 50  
|||||  
101 GAAAGAAAAAGAGACAGACAAAAAGAAAGAAATTAAGGCAATTGAAAAG 150  
50 ggluylsLysGluLysGluLysLysLysGluLysGluLysAspAsnL 67  
|||||  
151 GGAGAAAAAGAAAAAGAGCATTAAGAAATTAAGAAAAAGAAATTAATA 200  
67 ysgluysarggluysnlysvallysgluSerThrGlnLysGluLysGlu 83  
|||||  
201 AAGAAAAAGAGAAAAACAAAGTAAGAAATCCACACAGAAAGAAAAAGAA 250  
84 vallysgluGluLys 88  
|||||  
251 GTGAAGGAGAGAGAG 265

seq\_name: gb\_pat:A58694

## seq\_documentation\_block:

LOCUS A58694 137 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 13 from Patent WO9639505.  
ACCESSION A58694  
VERSION A58694.1 GI:3714252  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE 1 (bases 1 to 137)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 13 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES  
Location/Qualifiers  
source  
1..137  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 76 a 10 c 33 g 18 t  
ORIGIN

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x A58694 ..

Align seg 1/1 to: A58694 from: 1 to: 137

35 GluLysArgGluThrLysGluLysGluLys 45  
|||||



104 GAAAAAGACACACAAAAAGAAAAATTAAG 136

seq\_name: gb\_hcg:AC012577

seq\_documentation\_block:

LOCUS AC012577 164144 bp DNA linear HTG 30-MAR-2000

DEFINITION Homo sapiens clone RP11-134, WORKING DRAFT SEQUENCE, 47 unordered

pieces.

ACCESSION AC012577

VERSION AC012577.3 GI:7341804

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 164144)

Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Boucknight, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galligan, J., Garayna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,

Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Lien, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testa, S., Tjelle, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wymann, D., Ye, W. J., Zimmer, A., and Zody, M.

Direct Submission

Submitted (30-OCT-1999) Whitehead Institute/MIT Center for Genome

Research, 330 Charles Street, Cambridge, MA 02141, USA

On Mar 30, 2000 this sequence version replaced gi:6479098.

All repeats were identified using RepeatMasker:

Smit, A. P. A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W18R

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2500

Center clone name: L2500

----- Summary Statistics

Sequencing vector: MJ3; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136646 bases at least Q40  
Consensus quality: 147222 bases at least Q30  
Consensus quality: 152549 bases at least Q20  
Insert size: 173000; agarose-ff  
Insert size: 159544; sum-of-coverage  
Quality coverage: 3.1 in Q20 bases; agarose-ff  
Quality coverage: 3.3 in Q20 bases; sum-of-coverage

2568 4182: contig of 1615 bp in length  
4183 4282: gap of 100 bp  
4283 5706: contig of 1424 bp in length  
5707 5806: gap of 100 bp  
5807 7051: contig of 1245 bp in length  
7052 7151: gap of 100 bp  
7152 8775: contig of 1624 bp in length  
8776 8875: gap of 100 bp  
8876 10671: contig of 1796 bp in length  
10672 10771: gap of 100 bp  
10772 12406: contig of 1635 bp in length  
12407 12506: gap of 100 bp  
12507 13970: contig of 1464 bp in length  
13971 14070: gap of 100 bp  
14071 15920: contig of 1850 bp in length  
15921 16020: gap of 100 bp  
16021 18236: contig of 2216 bp in length  
18237 18336: gap of 100 bp  
18337 19315: contig of 1479 bp in length  
19316 19315: gap of 100 bp  
19316 21938: contig of 2023 bp in length  
21938 22038: gap of 100 bp  
22039 23976: contig of 1938 bp in length  
23977 24076: gap of 100 bp  
24077 25970: contig of 1894 bp in length  
25971 26070: gap of 100 bp  
26071 27824: contig of 1754 bp in length  
27825 27924: gap of 100 bp  
27925 30049: contig of 2125 bp in length  
30050 30149: gap of 100 bp  
30150 31707: contig of 1558 bp in length  
31708 31807: gap of 100 bp  
31808 33988: contig of 2181 bp in length  
33989 34088: gap of 100 bp  
34089 36181: contig of 2093 bp in length  
36182 36281: gap of 100 bp  
36282 39576: contig of 3295 bp in length  
39577 39676: gap of 100 bp  
39677 42561: contig of 2885 bp in length  
42562 42661: gap of 100 bp  
42662 45123: contig of 2462 bp in length  
45124 45223: gap of 100 bp  
45224 48511: contig of 3288 bp in length  
48512 48611: gap of 100 bp  
48612 51312: contig of 2701 bp in length  
51313 51412: gap of 100 bp  
51413 54343: contig of 2931 bp in length  
54344 54443: gap of 100 bp  
54444 57005: contig of 2562 bp in length  
57006 57105: gap of 100 bp  
57106 60893: contig of 3788 bp in length  
60894 60993: gap of 100 bp  
60994 63781: contig of 2788 bp in length  
63782 63881: gap of 100 bp  
63882 66724: contig of 2843 bp in length  
66725 66824: gap of 100 bp  
66825 69259: contig of 2435 bp in length  
69260 69359: gap of 100 bp  
69360 72510: contig of 3151 bp in length  
72511 72610: gap of 100 bp  
72611 76509: contig of 3899 bp in length  
76510 76609: gap of 100 bp  
76610 80140: contig of 3531 bp in length  
80141 80240: gap of 100 bp  
80241 84178: contig of 3938 bp in length  
84179 84278: gap of 100 bp  
84279 89364: contig of 5086 bp in length  
89365 89464: gap of 100 bp  
89465 94500: contig of 5036 bp in length  
94501 94600: gap of 100 bp  
94601 99471: contig of 4871 bp in length  
99472 99571: gap of 100 bp  
99572 104300: contig of 4729 bp in length



\* as soon as it is available and the accession number will  
\* be preserved.  
1 1828: contig of 1828 bp in length  
\* 1829 1928: gap of 100 bp  
\* 1929 26417: contig of 24489 bp in length  
\* 26418 26517: gap of 100 bp  
\* 26518 54437: contig of 27920 bp in length  
\* 54438 54537: gap of 100 bp  
\* 54538 101424: contig of 46887 bp in length  
\* 101425 101524: gap of 100 bp  
\* 101525 177957: contig of 76433 bp in length.  
Location/Qualifiers  
1..177957  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RP11-2N21"  
/clone\_lib="RP11-11 Human Male BAC"  
1..1828  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
1929..26417  
/note="assembly\_fragment"  
26518..54437  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
54538..101424  
/note="assembly\_fragment"  
101525..177957  
/note="assembly\_fragment"  
BASF COUNT 53204 a 34244 c 35114 g 54955 t 400 others  
ORIGIN  
alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
Alignment\_block:  
US-08-973-363-11 x AC025298 ..  
Align seg 1/1 to: AC025298 from: 1 to: 177957  
66 AsnlySGlUyArGluAsnlySVallySGlu 76  
|||||  
53821 AATAAGAGAAAAGCAGAAATTAAGTCAAGCA 53853  
seq\_name: gb\_pr:AC092372  
seq\_documentation\_block:  
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
ACCESSION AC092372  
VERSION AC092372.3 GI:17402768  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
2 (bases 1 to 101220)  
REFERENCE 2 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 101220)  
REFERENCE 3 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission

JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Fishing completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated total number of errors is 0.  
NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146,7kb). It is clipped at the overlap with AC012624.  
The number of bases overlapped is 90404.  
Location/Qualifiers  
1..101220  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone\_lib="RP11-58M12"  
BASE COUNT 34122 a 18862 c 17827 g 30409 t  
ORIGIN  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
Alignment\_block:  
US-08-973-363-11 x AC092372/rev ..  
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220  
67 LysGluLyArGluAsnlySVallySGlu 76  
|||||  
25212 AAAGAAAAGAGAAACAAAGTAAAGAA 25183  
seq\_name: gb\_pr:AC012624  
seq\_documentation\_block:  
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001  
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.  
ACCESSION AC012624  
VERSION AC012624.6 GI:14993679  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
2 (bases 1 to 134365)  
REFERENCE 2 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 134365)  
REFERENCE 3 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 134365)  
REFERENCE 4 (bases 1 to 134365)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.  
Location/Qualifiers  
1..134365  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"

BASE COUNT 40414 a 24497 c 25503 g 43951 t  
ORIGIN /clone="CTD-2082117"

alignment\_scores: 10.00 Length: 10  
Quality: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

67 LysgilyuAAGGluAAsnlyValysgu 76  
|||||  
119971 AAGGAAAAAGGAGAAACAACTAAAAAGCA 120000

seq\_name: gb\_hlg:AC021449

seq\_documentation\_block:  
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000  
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
pieces.  
AC021449  
VERSION AC021449.3 GI:10047806  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-58M12  
Unpublished  
2 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,  
Cooper,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeArliano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lander,T., Lehoczeky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5154  
Center clone name: 58\_M\_12  
----- Summary Statistics  
Sequencing vector: MJ3; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 144000; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 38820: contig of 38820 bp in length  
38821 38920: gap of 100 bp  
38921 40411: contig of 1491 bp in length  
40412 40511: gap of 100 bp  
40512 43279: contig of 2768 bp in length  
43280 43379: gap of 100 bp  
43380 46905: contig of 3526 bp in length  
46906 47005: gap of 100 bp  
47006 51830: contig of 4825 bp in length  
51831 51930: gap of 100 bp  
51931 62619: contig of 10689 bp in length  
62620 62719: gap of 100 bp  
62720 75408: contig of 12689 bp in length  
75409 75508: gap of 100 bp  
75509 92516: contig of 17008 bp in length  
92517 92616: gap of 100 bp  
92617 106409: contig of 13793 bp in length  
106410 106509: gap of 100 bp  
106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_1lb="RPCT-11 Human Male BAC"  
1. 38820  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
38921. 40411  
/note="assembly\_fragment"  
40512. 43279  
/note="assembly\_fragment"  
43380. 46905  
/note="assembly\_fragment"  
47006. 51830  
/note="assembly\_fragment"  
51931. 62619  
/note="assembly\_fragment"  
62720. 75408  
/note="assembly\_fragment"  
75509. 92516  
/note="assembly\_fragment"  
92617. 106409  
/note="assembly\_fragment"  
106510. 143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

alignment\_scores: 10.00 Length: 10  
Quality: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:
US-08-973-363-11 x AC021449 ..
Align seg 1/1 to: AC021449 from: 1 to: 143079
67 LysGluYsArGluAsnLysValLysGlu 76
|||||
117867 AAAGAAAAAGAGAAACAAAGTAAAGAA 117896

seq_name: gb_htg:AC008531

seq_documentation_block:
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,
7 ordered pieces.
ACCESSION AC008531.3 GI:12830078
VERSION AC008531
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 145659)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
2 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 14, 2001 this sequence version replaced gi:1528342.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Estimated coverage: 6.27 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 56174: contig of 56174 bp in length
56175 56174: gap of unknown length
56275 100874: contig of 44600 bp in length
100875 100974: gap of unknown length
100975 113127: contig of 12153 bp in length
113128 113227: gap of unknown length
113228 118190: contig of 4963 bp in length
118191 118290: gap of unknown length
118291 119694: contig of 1404 bp in length
119695 119794: gap of unknown length
119795 123297: contig of 3503 bp in length
123298 123398: gap of unknown length
123398 145659: contig of 22262 bp in length.
Location/Qualifiers

```

```

source
1. .145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-480B11"
/clone.lib="Caltech human BAC library C"
BASE COUNT 42561 a 26305 c 27580 g 48609 t 600 others
ORIGIN

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AC008531 ..
Align seg 1/1 to: AC008531 from: 1 to: 145659
67 LysGluYsArGluAsnLysValLysGlu 76
|||||
48484 AAAGAAAAAGAGAAACAAAGTAAAGAA 48513

seq_name: gb_htg:AC091946

seq_documentation_block:
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 193446)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE Unpublished
2 (bases 1 to 193446)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RPCT-11_36012
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-gel estimation
Quality coverage: 5.38 in Q20 bases; agarose-gel estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1358: contig of 1358 bp in length
1359 1458: gap of unknown length

```

```
* 1459 2667: contig of 1209 bp in length
* 2668 2767: gap of unknown length
* 3823 3823: contig of 1056 bp in length
* 3824 3923: gap of unknown length
* 3924 5920: contig of 1997 bp in length
* 5921 6020: gap of unknown length
* 6021 7494: contig of 1474 bp in length
* 7495 7595: gap of unknown length
* 7595 9097: contig of 1502 bp in length
* 9097 9197: gap of unknown length
* 9197 10261: contig of 1064 bp in length
* 10261 10360: gap of unknown length
* 10361 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length
* 12561 14611: contig of 2051 bp in length
* 14612 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 16481: gap of unknown length
* 16482 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31088: gap of unknown length
* 31089 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 51745: gap of unknown length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74133: gap of unknown length
* 74134 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169210: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
location/Qualifiers
  1..193446
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="RP11-36012"
    /clone_idb="RPC1 human BAC library 11"
  57571 a 35252 c 37061 g 60360 t 3202 others
```

FEATURES  
SOURCE

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AC091946 ..  
Align seg 1/1 to: AC091946 from: 1 to: 193446

67 LysGluLysArgGluAsnLysValLysGlu 76

89378 AAGAAAGAGAGAGAAACCACTAAAGCA 89407

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778

VERSION AC026778.4 GI:14277282

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 2 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 3 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 4 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 5 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 6 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 7 (bases 1 to 195433)

TITLE Direct Submission

JOURNAL DOE Joint Genome Institute.

REFERENCE 8 (bases 1 to 195433)

TITLE Direct Submission

```
|||||
28018 AAGAGAAAAAGGAAACAAAGTAAAGAA 27989
seq_name: gb_pr:AC022121

seq_documentation_block:
  LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001
  DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
  ACCESSION AC022121
  VERSION AC022121.6 GI:15375145
  KEYWORDS HTG.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE
    1 (bases 1 to 219258)
    DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL
    Title Unpublished
    Reference 2 (bases 1 to 219258)
    DOE Joint Genome Institute.
  JOURNAL
    Title Direct Submission
    Reference Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
    Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  JOURNAL
    Title Direct Submission
    Reference 3 (bases 1 to 219258)
    DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL
    Title Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
    Drive, Walnut Creek, CA 94598, USA
  JOURNAL
    Title Direct Submission
    Reference DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL
    Title Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
    Drive, Walnut Creek, CA 94598, USA
  JOURNAL
    Title On Aug 30, 2001 this sequence version replaced gi:15148108.
  COMMENT
    Draft Sequence Produced by DOE Joint Genome Institute
    www.jgi.doe.gov
    Flushing Completed at Stanford Human Genome Center
    www.shgc.stanford.edu
    Quality: Phrap Quality >=40 99.8% of Sequence;
    Estimated Total Number of Errors is 0.4.
    STS Content:
    WI-5811 G04974
    WI-13675 G23101
    SHGC-58345 G38487
    SHGC-103595 G57841.
    Location/Qualifiers
    FEATURES
      source
        1..219258
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="CTD-2007H13"
    BASE COUNT 71954 a 42062 c 40933 g 64309 t
    ORIGIN
      alignment_scores:
        Quality: 10.00 Length: 10
        Ratio: 1.000 Gaps: 0
        Percent Similarity: 100.000 Percent Identity: 100.000
      alignment_block:
        US-08-973-363-11 x AC022121/rev ..
      Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258
      67 LyagLUySARGLuAaNLySValySGLu 76
      ||||||||||||||||||||||||||||
      87726 AAGAGAAAAAGGAAACAAAGTAAAGAA 87697
      seq_name: gb_htg:AC092382
      seq_documentation_block:
        LOCUS AC092382 276181 bp DNA linear HTG 03-JUL-2001
        DEFINITION Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE,
        47 overlapping pieces.
        ACCESSION AC092382
        VERSION AC092382.1 GI:14589571
        KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
        SOURCE human.
        ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
        REFERENCE
          1 (bases 1 to 276181)
          DOE Joint Genome Institute.
        JOURNAL
          Title Unpublished
          Reference 2 (bases 1 to 276181)
          DOE Joint Genome Institute.
        JOURNAL
          Title Direct Submission
          Reference Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
          Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
        JOURNAL
          Title Center: Joint Genome Institute
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          Project Information
          Center Project Name: 435334
          Center clone name: RPC1-11_75H1
          -----
          Summary Statistics
          Consensus quality: 229677 bases at least Q40
          Consensus quality: 256163 bases at least Q30
          Consensus quality: 260799 bases at least Q20
          Estimated insert size: 174820; agarose-fp estimation
          Estimated insert size: 271581; sum-of-contigs estimation
          Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
          Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 47 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          1
          1062 1061: contig of 1061 bp in length
          1161: gap of unknown length
          1162 2827: contig of 1666 bp in length
          2828 4327: gap of unknown length
          4328 4327: contig of 1300 bp in length
          4329 5718: gap of unknown length
          5719 6983: contig of 1251 bp in length
          6984 7083: gap of unknown length
          7084 8422: contig of 1339 bp in length
          8423 8522: gap of unknown length
          8523 9771: contig of 1249 bp in length
          9772 9871: gap of unknown length
          9872 11198: contig of 1327 bp in length
          11199 11298: gap of unknown length
          11299 12514: contig of 1116 bp in length
          12515 13806: gap of unknown length
          13807 13906: contig of 1292 bp in length
          13907 15360: gap of unknown length
          15361 15460: contig of 1454 bp in length
          15461 16667: gap of unknown length
          16668 16767: contig of 1207 bp in length
          16768 18390: gap of unknown length
          18391 18490: contig of 1623 bp in length
          18491 19541: gap of unknown length
          19542 19541: contig of 1051 bp in length
          19543 21156: gap of unknown length
          21156 21156: contig of 1515 bp in length
```











```
1 AspGluIleValSerValIysHisLeuHisLysLysLysIleLysThrGlu.L 17
|||||
1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATTAAGAAAAAGAAAA 50
17 ySGluAsnGluGluLysProGluProAspIleGlyIleLysLysGluAla 33
|||||
51 AAGAAAAATGAAGAAAAAGCCTGAGCCAGATTGTATTAAGAAAGAAAGCT 100
34 GluGluLysArgGluThrLysGluLysGluAsnLysArgGluLeuLysAr 50
|||||
101 GAAGAAAAAAGAGACAAAAAGAAAGAAATTAAGGAATTGAAAAAG 150
50 gGluLysLysGluLysGluAspLysLysGluLeuLysGluLysAspAsnL 67
|||||
151 GGAGAAAAAAGAAAAAGAGATTAAGAAATTAAGAAAGAAATTAATA 200
67 ySGluLysArgGluAsnLysValLysGluSerThrGluLysGluLysGlu 83
|||||
201 AAGAAAAAGAGAAAAAGAAAGTAAAGATCCACACAGAAAAAGAAAAAGAA 250
84 ValLysGluGluLys 88
|||||
251 GTGAAGGAAGAGAAC 265
```

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142753

seq\_documentation\_block:

ID AA142753 standard; CDNA; 137 BP.

AA142753:

12-MAR-1997 (first entry)

Chicken CHD-W clone CC14 3' motif.

Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

Gallus sp.

Location/Qualifiers

misc\_difference 52

/\*tag= a

/note= "base 52 disrupts the reading frame for

the translated amino acid sequence given

in Fig 7"

W09639505-A1.

12-DEC-1996.

05-JUN-1996; 96WC-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

P-PSDB; AAW08145.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Disclosure; Fig 7; 76pp; English.

A composite (incomplete) nucleotide sequence (AA142753) and putative

translation (AAW08145) sequence are provided of a motif that is found

CC spliced to the 3' end of chicken CHD-W clone CC14 (see also AA142754).

CC There are no splice donor or acceptor sites within the motif

CC suggesting it is a final rather than an intermediary product of

CC splicing. The motif is also found as an insert in some CHD-1A clones

CC (see also AA142752).

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

alignment\_scores:

Quality: 11.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AA142753

Align seg 1/1 to: AA142753 from: 1 to: 137

35 GluLysArgGluThrLysGluLysGluAsnLys 45

|||||

104 GAAAAAAGAGAGACAAAGAAAGAAATTAAG 136

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA142753

seq\_documentation\_block:

ID AAC04835 standard; CDNA; 152 BP.

AAC04835:

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 8910.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for

obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 8910; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. No ORF has yet been conclusively

identified within the present sequence. The 5' ESTs were prepared from

total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

sequences usually correspond mainly to the 3' untranslated region (UTR)

of the mRNA because they are often obtained from oligo-dT primed cDNA

libraries. Such ESTs are not well suited for isolating cDNA sequences

derived from the 5' ends of mRNAs and even in those cases where longer

cDNA sequences have been obtained, the full 5' UTR is rarely included.

5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

in diagnostic, forensic, gene therapy and chromosome mapping procedures.

They are used to obtain upstream regulatory sequences and to design

expression and secretion vectors.

Sequence 152 BP; 37 A; 33 C; 22 G; 60 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AAC04835/rev ..

Align seg 1/1 to reverse of: AAC04835 from: 1 to: 152

52 LysylsGluLysGluAspLysLys 59  
|||||  
107 AAAAAGAGAGAGAGAGAGAGAA 84

seq\_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH70996

seq\_documentation\_block:  
ID: AAH70996 standard: cDNA: 271 BP.

AAH70996:

19-SEP-2001 (first entry)

Human cervical cancer marker nucleic acid 2270.

Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

Homo sapiens.

MO200142467-A2.

14-JUN-2001.

08-DEC-2000; 2000WO-US33312.

08-DEC-1999; 99US-0169681.

21-DEC-1999; 99US-0171350.

14-MAR-2000; 2000US-0189315.

12-MAY-2000; 2000US-0203791.

09-JUN-2000; 2000US-0210600.

21-JUL-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X;

WPI; 2001-375006/39.

New isolated nucleic acid for diagnosing and treating cervical cancer

and for assessing and detecting compounds for treating the cancer -

Claim 1; Page 480; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with

cervical cancer with cytostatic activity. The nucleic acids and encoded

polypeptides are useful: to assess if a patient is afflicted with

cervical cancer or has a pre-malignant condition; to monitor the

progression of cervical cancer or a premalignant condition in a patient;

and to select and/or assess the efficacy of a compound or therapy for

inhibiting cervical cancer in a patient. The nucleic acids may also be

useful for gene therapy.

Sequence 271 BP; 98 A; 56 C; 75 G; 41 T; 1 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AAH70996 ..

Align seg 1/1 to: AAH70996 from: 1 to: 271

20 GluLysProGluProAspIle 27  
|||||  
128 GAGAGAGAGAGAGAGAGAGAT 151

seq\_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA74794

seq\_documentation\_block:  
ID: ABA74794 standard: DNA: 287 BP.

ABA74794:

01-FEB-2002 (first entry)

Human foetal liver single exon nucleic acid probe #23099.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

Homo sapiens.

MO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00669.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human foetal liver -

Claim 4; SEQ ID NO 23099; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for

measuring human gene expression in a sample derived from human foetal

liver. The single exon nucleic acid probes may be used for predicting,

measuring and displaying gene expression in samples derived from human

foetal liver. The present sequence is a single exon nucleic acid

probe of the invention.

Note: The sequence data for this patent did not form part of the

printed specification, but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x ABA74794 ..

Align seg 1/1 to: ABA74794 from: 1 to: 287

49 LysArgGluLysGluLysGlu 56  
|||||  
16 AAAAGGAGAGAGAGAGAGAG 39

```
seq_name: /SIDSL/gcdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.ABA39495
seq_documentation_block:
ID ABA39495 standard; DNA: 287 BP.
XX
AC ABA39495;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #17961 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN W0200157274-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
PS Claim 4; SEQ ID NO 17961; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x ABA39495 ..
Align seg 1/1 to: ABA39495 from: 1 to: 287

49 LysARGlULySLySGlULySGlu 56
|||||
16 AAAAGGAGAGAGAGAGAGAGAG 39
```

```
seq_name: /SIDSL/gcdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.AAK23286
seq_documentation_block:
ID AAK23286 standard; DNA: 287 BP.
XX
AC AAK23286;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 23277.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN W0200157275-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
PS Example 4; SEQ ID NO: 23277; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AAK23286 ..
Align seg 1/1 to: AAK23286 from: 1 to: 287

49 LysARGlULySLySGlULySGlu 56
|||||
16 AAAAGGAGAGAGAGAGAGAGAG 39

seq_name: /SIDSL/gcdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.AAK49450
seq_documentation_block:
ID AAK49450 standard; DNA: 287 BP.
XX
AC AAK49450;
```

```
XX 06-NOV-2001 (first entry)
DT
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 24007.
XX
XX Human bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 24007; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 8.00      Length: 8
XX      Ratio: 1.000      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX      US-08-973-363-11 x AAK49450 ..
XX
XX Align seg 1/1 to: AAK49450 from: 1 to: 287
XX
XX      49 LysATGtLulysLysGtLulysGlu 56
XX      |||||
XX      16 AAAAGGAGAGAGAGAGAGAGAGAG 39
XX
XX seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI26561
XX
XX seq_documentation_block:
XX ID AAI26561 standard; DNA; 287 BP.
XX
XX AC AAI26561;
XX
XX DT 12-OCT-2001 (first entry)
XX
XX DE Probe #16494 for gene expression analysis in human cervical cell sample.
XX
XX KM Probe: human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
```

```
OS Homo sapiens.
XX
XX W0200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 16494; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC
XX
XX Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 8.00      Length: 8
XX      Ratio: 1.000      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX      US-08-973-363-11 x AAI26561 ..
XX
XX Align seg 1/1 to: AAI26561 from: 1 to: 287
XX
XX      49 LysATGtLulysLysGtLulysGlu 56
XX      |||||
XX      16 AAAAGGAGAGAGAGAGAGAGAGAG 39
XX
XX seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI55319
XX
XX seq_documentation_block:
XX ID AAI55319 standard; DNA; 287 BP.
XX
XX AC AAI55319;
XX
XX DT 17-OCT-2001 (first entry)
XX
XX DE Probe #24005 used to measure gene expression in human placenta sample.
XX
XX KM Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX W0200157272-A2.
XX
```

```
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 24005; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AA155319 ..

Align seg 1/1 to: AA155319 from: 1 to: 287

49 LysArgGluLysLysGluLysGlu 56
   |||||
16 AAAAGGAGAAAGCAAGAGAGAG 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH72066
seq_documentation_block:
ID AAH72066 standard; cDNA: 314 BP.
XX
XX AAH72066;
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 3340.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US33312.
XX
XX 08-DEC-1999; 99US-0169681.
XX 21-DEC-1999; 99US-0171350.
XX 14-MAR-2000; 2000US-0189315.
XX 12-MAY-2000; 2000US-0203791.
XX 09-JUN-2000; 2000US-0210600.
XX 21-JUL-2000; 2000US-0220114.
```

```
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX and for assessing and detecting compounds for treating the cancer -
XX
XX Claim 1; Page 649; 1051pp; English.
XX
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids and encoded
XX polypeptides are useful: to assess if a patient is afflicted with
XX cervical cancer or has a pre-malignant condition; to monitor the
XX progression of cervical cancer or a premalignant condition in a patient;
XX and to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.
XX
XX Sequence 314 BP; 124 A; 58 C; 82 G; 50 T; 0 other;

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AAH72066 ..

Align seg 1/1 to: AAH72066 from: 1 to: 314

20 GluGluLysProGluProAspIle 27
   |||||
104 GAAGAAAGCCAGACAGACATA 127

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH69191
seq_documentation_block:
ID AAH69191 standard; cDNA: 357 BP.
XX
XX AAH69191;
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 465.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US33312.
XX
XX 08-DEC-1999; 99US-0169681.
XX 21-DEC-1999; 99US-0171350.
XX 14-MAR-2000; 2000US-0189315.
XX 12-MAY-2000; 2000US-0203791.
XX 09-JUN-2000; 2000US-0210600.
XX 21-JUL-2000; 2000US-0220114.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX
```



PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1: Page 186; 1051pp; English.  
XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 357 BP; 125 A; 73 C; 92 G; 59 T; 8 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-11 x AAH69191 ..  
  
Align seg 1/1 to: AAH69191 from: 1 to: 357  
  
20 GluGluLysProGluProaspIle 27  
|||||  
105 GAAAGAAAGCCAGAACGACATG 128  
  
seq\_name: /SIDSI/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA62258  
seq\_documentation\_block:  
ID ABA62258 standard; DNA; 575 BP.  
XX  
AC ABA62258;  
XX  
XX 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #10563.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe: ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015727-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001MO-US000669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver -  
XX  
XX Claim 1: SEQ ID NO 10563; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-11 x ABA62258 ..  
  
Align seg 1/1 to: ABA62258 from: 1 to: 575  
  
49 LysArgGluLysGluLysGlu 56  
|||||  
69 AAAAGCAGACAGACAGACAGAG 92  
  
seq\_name: /SIDSI/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA29604  
seq\_documentation\_block:  
ID ABA29604 standard; DNA; 575 BP.  
XX  
AC ABA29604;  
XX  
XX 23-JAN-2002 (first entry)  
XX  
DE Probe #8070 for gene expression analysis in human heart cell sample.  
XX  
XX  
XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015727-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001MO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX Claim 1: SEQ ID NO 8070; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x ABA29604 ..

Align seg 1/1 to: ABA29604 from: 1 to: 575

49 LysArgGluLysGluLysGlu 56  
|||||  
69 AAAAGGAGAGAGAGAGAGAG 92

seq\_name: /SIDSL/gcgdata/hold\_geneseq/geneseqn-emb1/NA2001A.DAT.AAK10591

seq\_documentation\_block:

ID AAK10591 standard; DNA; 575 BP.

AC AAK10591:

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 10582.

KW Human: brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -

PS Example 4; SEQ ID NO: 10582; 650bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AAK10591 ..


Align seg 1/1 to: AAK10591 from: 1 to: 575

49 LysArgGluLysGluLysGlu 56  
|||||  
69 AAAAGGAGAGAGAGAGAGAG 92

Mon Aug 5 11:51:38 2002

us-08-973-363-11.olip2n.rng

Page 9



**THIS PAGE BLANK (USPTO)**

Page 1

Sequence	Strid	Orf	Score	Len	Documentation
gb_estc1:BM015557	+	10.00	159.44	12.37	677
gb_estc1:AM053734	+	9.00	149.02	47.07	290
gb_estc1:BC661118	+	9.00	146.60	64.17	413
gb_estc1:BC510863	+	9.00	145.73	68.78	447
gb_gsts:A2257993	+	9.00	145.73	71.73	469
gb_estc1:AM050334	+	9.00	145.15	77.20	510
gb_estc1:AM703096	+	9.00	145.05	78.26	518
gb_gsts:A2879168	+	9.00	142.16	75.81	542
gb_gsts:BC738481	+	9.00	142.16	113.43	791
gb_estc1:BC653397	+	9.00	141.94	116.67	817
gb_estc1:BC621310	+	9.00	141.94	116.67	817
gb_estc1:BP161157	+	9.00	141.22	127.99	1140
gb_estc1:BC531075	+	9.00	139.67	156.23	1140
gb_estc1:BC653128	+	9.00	139.62	157.37	1149
gb_gsts:ACG178168	+	9.00	138.71	176.41	1310
gb_gsts:A0286520	+	8.00	139.19	166.69	114
gb_estc1:AM856568	+	8.00	138.03	192.61	135
gb_estc1:AT128483	+	8.00	137.98	195.11	137
gb_estc1:R1491709	+	8.00	137.88	195.11	138
gb_estc1:BV11941975	+	8.00	135.84	326.93	186
gb_gsts:A2583184	+	8.00	135.42	329.43	198
gb_estc1:AM020742	+	8.00	135.01	283.63	210
gb_estc1:AA393117	+	8.00	134.45	304.83	228
gb_estc1:BB601962	+	8.00	134.45	304.83	228
gb_estc1:BB601955	+	8.00	133.66	334.25	230
gb_estc1:BC621310	+	8.00	133.66	335.47	236
gb_estc1:BC174143	+	8.00	132.93	347.75	263
gb_estc1:BB009657	+	8.00	132.93	347.75	263
gb_estc1:BB009657	+	8.00	132.98	379.03	287
gb_estc1:BB009657	+	8.00	132.74	379.85	293
gb_gsts:A0584452	+	8.00	132.65	384.39	297
gb_estc1:BB009657	+	8.00	132.65	384.39	297
gb_estc1:BB009657	+	8.00	132.60	386.66	299
gb_estc1:BB009657	+	8.00	132.58	387.80	300
gb_estc1:BB009657	+	8.00	132.55	388.93	301
gb_estc1:BB406895	+	8.00	132.42	395.71	307
gb_estc1:BB406895	+	8.00	132.42	395.71	307
gb_estc1:BB694264	+	8.00	132.38	430.97	309
gb_estc1:BB059258	+	8.00	132.33	400.23	311
gb_estc1:BM059264	+	8.00	132.33	400.23	311
gb_estc1:BM015557	+	10.00	159.44	12.37	677
gb_estc1:AM053734	+	9.00	149.02	47.07	290
gb_estc1:BC661118	+	9.00	146.60	64.17	413
gb_estc1:BC510863	+	9.00	145.73	68.78	447
gb_gsts:A2257993	+	9.00	145.73	71.73	469
gb_estc1:AM050334	+	9.00	145.15	77.20	510
gb_estc1:AM703096	+	9.00	145.05	78.26	518
gb_gsts:A2879168	+	9.00	142.16	75.81	542
gb_gsts:BC738481	+	9.00	142.16	113.43	791
gb_estc1:BC653397	+	9.00	141.94	116.67	817
gb_estc1:BC621310	+	9.00	141.94	116.67	817
gb_estc1:BP161157	+	9.00	141.22	127.99	1140
gb_estc1:BC531075	+	9.00	139.67	156.23	1140
gb_estc1:BC653128	+	9.00	139.62	157.37	1149
gb_gsts:ACG178168	+	9.00	138.71	176.41	1310
gb_gsts:A0286520	+	8.00	139.19	166.69	114
gb_estc1:AM856568	+	8.00	138.03	192.61	135
gb_estc1:AT128483	+	8.00	137.98	195.11	137
gb_estc1:R1491709	+	8.00	137.88	195.11	138
gb_estc1:BV11941975	+	8.00	135.84	326.93	186
gb_gsts:A2583184	+	8.00	135.42	329.43	198
gb_estc1:AM020742	+	8.00	135.01	283.63	210
gb_estc1:AA393117	+	8.00			

```

gb_estl:AWI44996 + 8.00 133.03 415.98 335 | AMI44996 rs44d08.y1 Sommer P
gb_estl:DJ3021 + 8.00 131.86 424.94 333 | DJ3021 CELK11268F.vuji Kohar
gb_estl:A1551690 - - 8.00 131.76 430.52 338 | A1551690 v66b0d4.y1 Knowles
gb_estl:BE247224 + 8.00 131.72 432.75 340 | BE247224 TCBAPIE3832 Pediatr
gb_estl:B1038399 - 8.00 131.68 434.98 342 | B1038399 RCS-NT0266-180101-0
gb_gss:AO602323 + 8.00 131.66 434.98 342 | AO602323 HS-2114.A2 AlO TTC
gb_estl:AI197098 - 8.00 131.66 436.10 343 | AI197098 uds7d08.r1 Soares_N

seq_name: gb_estl2:BM015597

seq_documentation_block:
LOCUS BM015597 677 bp mRNA linear EST_30-OCT-2001
DEFINITION G03641828F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5417887.5,
ACCESSION BM015597
VERSION BM015597.1 GI:16529951
KEYWORDS Est.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs@femail.nih.gov
Tissue Procurement: DCD/DTP
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM2066.row:g column: 08
High quality sequence stop: 674.
FEATURES
Source Location/Qualifiers
1..677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image=5417887
/clone_lib="NIH-MGC_87"
/lusue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectional; Oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT 305 a 92 c 153 g 127 t
ORIGIN
alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-J1 x BM015597 ..
Align seg 1/1 to: BM015597 from: 1 to: 677
67 TgcGtUgArGcUAsNtYsValYsgU 76
|||||
288 AAAAGAAAAAGAAACAAGTAAGA 317

seq_name: gb_estl1:A0053734

seq_documentation_block:
LOCUS A0053734 290 bp mRNA linear EST_28-APR-1999
DEFINITION Dictyostelium discoideum SL (H.Urisuthara) Dictyostelium
discoideum cDNA clone SIdj576. mRNA sequence.
```

ACCESSION AU053734  
VERSION AU053734.1 GI:4702216  
KEYWORDS  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
REFERENCE Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
AUTHORS Mori, T., Urushihara, H., Salto, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Mitra, B. N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
TITLE Developmental cDNA in Dictyostelium discoideum  
JOURNAL Unpublished (1998)  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan  
Email: d402nuesakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium discoideum cDNA project in Japan.  
FEATURES  
source 1..290  
location/Qualifiers  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLJ576"  
/clone\_lib="Dictyostelium discoideum SL (H. Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 158 a 28 c 55 g 47 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x AU053734 ..  
Align seg 1/1 to: AU053734 from: 1 to: 290  
seq\_name: gb\_est2:BG611198  
31 LysGluAlaGluGluLysArgGluThr 39  
|||||  
2 AAAGAAGCTGGAAGAAAGAGAGAACT 28  
seq\_name: gb\_est2:BG611198  
seq\_documentation\_block:  
LOCUS BG611198 413 bp mRNA linear EST 18-APR-2001  
DEFINITION 60261217F1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4737522 5',  
mRNA sequence.  
ACCESSION BG611198  
VERSION BG611198.1 GI:13662569  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE NIH-MGC http://mgi.nci.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: DCTMD/DTF  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM1605 row: j column: 19  
High quality sequence stop: 230.  
Location/Qualifiers

source 1..413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4737522"  
/clone\_lib="NIH\_MGC\_60"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
site\_1: SfiI (ggccattggcc); site\_2: SfiI (ggccattggcc)  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGCGCGCGCAGTGT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
Library."  
BASE COUNT 176 a 77 c 104 g 56 t  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x BG611198 ..  
Align seg 1/1 to: BG611198 from: 1 to: 413  
seq\_name: gb\_est2:BI508663  
52 LysLysGluLysGluAspLysGlu 60  
|||||  
326 AAGAAGGAAAGAAAGACAAAGAGAG 352  
seq\_name: gb\_est2:BI508663  
seq\_documentation\_block:  
LOCUS BI508663 447 bp mRNA linear EST 29-AUG-2001  
DEFINITION BB170001B10E08 Bee Brain Normalized/Subtracted Library, BB17 Apis  
mellifera cDNA clone BB170001B10E08 5', mRNA sequence.  
ACCESSION BI508663  
VERSION BI508663.1 GI:15359037  
KEYWORDS EST.  
SOURCE honeybee.  
ORGANISM Apis mellifera  
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
AUTHORS Eukaryota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata  
; Apoidea; Apidae; Apis.  
TITLE Whitfield, C.W., Soares, B., Robertson, H.M., Pardinas, J., Liu, L.,  
Smoller, D. and Robinson, G.E.  
JOURNAL An Expressed Sequence Tag Resource for Studies of Brain and  
Behavior in the Honey Bee  
COMMENT Unpublished (2001)  
Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu  
This research was funded by the University of Illinois Critical  
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation  
Award in Functional Genomics to G.E. Robinson and an NSF  
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
PCR Primers  
FORWARD: TAATACGACTCACTATAGG  
BACKWARD: ATTACCTCCTACTTAAG  
Insert Length: 447 Std Error: 0.00  
Plate: BB170001B10 row: E column: 08

Seq primer: AGCGATACATTTTCACACGGA  
High quality sequence stop: 447.  
Location/Qualifiers  
source  
1. 447  
/organism="Apis mellifera"  
/strain="mixed strains of European bees, predominantly  
A.m. ligustica"  
/db\_xref="taxon:7460"  
/clone="BBI70001B10P08"  
/clone.lib="Bee Brain Normalized/Subtracted Library, BBI7"  
/sex="female"  
/tissue.type="brain"  
/dev\_stage="adult worker honey bee"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pT73-Pac; Site:1: EcoRI;  
Site:2: NotI. This BBI7 cDNA library was generated by  
subtraction of the BBI6 library with 4000 previously  
sequenced clones. The BBI7 library was contributed by the  
Soares laboratory and it was constructed and normalized  
as described by Bonaldo, M.F., Lennon, G. and Soares,  
M.B. (1996), Genome Research 6(9): 791-806. RNA was  
prepared from dissected brains of adult worker bees of  
various ages and various behavioral groups."

BASE COUNT 127 a 94 c 70 g 156 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x B1508663/rev ..

Align seg 1/1 to reverse of: B1508663 from: 1 to: 447

75 LysgluSerThrGlnLysGluLysGlu 83  
|||||  
149 AAGAGAAAGTACACAAAGAAAAAGAA 123

seq\_name: gb\_gss:A2257993

seq\_documentation\_block:  
LOCUS A2257993 469 bp DNA linear GSS 26-JUL-2000  
DEFINITION RPECI-23-120M12.TV RPECI-23 Mus musculus genomic clone RPECI-23-120M12  
KEYWORDS  
ACCESSION A2257993  
VERSION A2257993.1 GI:9463121  
GSS.  
SOURCE  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 469)  
Zhao,S., Niernan,W., Feldblum,T., Malek,J., Shatsman,S., Aklnret  
B., Levins,M., McGann,S., Tsagaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from library RPECI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhaog@igf.org  
Clones are derived from the mouse BAC library RPECI-23. For BAC  
library availability, please contact Peter de Jong  
(peter@edj.med.buffalo.edu). Clones may be purchased from  
BACAC Resources (http://bacac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.igf.org/cdb/bac\_ends/mouse/bac\_end\_intro.html

TITLE  
JOURNAL  
COMMENT

Plate: 120 row: M column: 12  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
source  
1. 469  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPECI-23-120M12"  
/clone.lib="RPECI-23"  
/sex="female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site:1:  
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 144 a 93 c 85 g 147 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x A2257993 ..

Align seg 1/1 to: A2257993 from: 1 to: 469

38 GluThrLysGluLysGluLysGlu 46  
|||||  
203 GAACAAAGAGGAAAGACACAGAGC 229

seq\_name: gb\_est1:A0053344

seq\_documentation\_block:  
LOCUS A0053344 510 bp mRNA linear EST 28-APR-1999  
DEFINITION A0053344 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
discoideum cDNA clone SL1432, mRNA sequence.  
ACCESSION A0053344  
VERSION A0053344.1 GI:4701826  
KEYWORDS  
SOURCE EST.  
ORGANISM Dictyostelium discoideum.  
Dictyostelium discoideum.  
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
1 (bases 1 to 510)  
Mori,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
Yoshino,R., Maeda,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H.,  
Williams,J., and Tanaka,Y.  
Developmental cDNA in Dictyostelium discoideum  
Unpublished (1998)  
Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT - Dictyostelium discoideum cDNA project in Japan.  
Location/Qualifiers  
1. 510  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SL1432"  
/clone.lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"

BASE COUNT 263 a 64 c 63 g 120 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AU053344 ..

Align seg 1/1 to: AU053344 from: 1 to: 510

31 lysgluAlaGlulysArgGluThr 39  
|||||  
307 AAGGAAGCTGAGAAAGAGAGAACT 333

seq\_name: gb\_est1:AW790096

seq\_documentation\_block:

LOCUS AW790096 518 bp mRNA linear EST 01-MAY-2001  
DEFINITION C01520-F lambda Zap, StrataGene Blumeria graminis f. sp. hordei  
CDNA clone C01520 similar to 101 kd malaria antigen, mRNA sequence.

ACCESSION AW790096

VERSION AW790096.1 GI:13901693

KEYWORDS EST.

SOURCE Blumeria graminis f. sp. hordei.

ORGANISM Blumeria graminis f. sp. hordei

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
Erysiphales; Erysiphaceae; Blumeria.

REFERENCE 1 (bases 1 to 518)

Thomas, S.W., Rasmussen, S.W., Glaring, M.A., Rousier, J.A. and Oliver

AUTHORS

TITLE Gene identification in the fungal pathogen Blumeria graminis by

expressed sequence tag analysis

UNPUBLISHED (2000)

CONTACT: Rasmussen, S.W.

Department of Yeast Genetics

Carlsberg Laboratory

10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark

Tel: 45 3327 5230

Fax: 45 3327 4766

Email: swr@crc.dk

High quality sequence stop: 518

FEATURES

Location/Qualifiers

1..518

/organism="Blumeria graminis f. sp. hordei"

/db\_xref="taxon:62688"

/clone="C01520"

/cell\_type="conidia"

/lab\_host="Hordeum vulgare"

BASE COUNT 131 a 99 c 84 g 204 t

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AW790096/rev ..

Align seg 1/1 to reverse of: AW790096 from: 1 to: 518

50 ArgGluLysGluLysGluAspLys 58  
|||||

451 AAGGAAAGAGAAAGAGAGAGTAAG 425

seq\_name: gb\_gss:AZ879168

seq\_documentation\_block:

LOCUS AZ879168

DEFINITION RPCI-23-194A5. TJ RPCI-23 Mus musculus genomic clone RPCI-23-194A5,

DNA sequence.

ACCESSION AZ879168

VERSION AZ879168.1 GI:13197234

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 542)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret,

B., Levins, M., Megam, S., Ysega, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

CONTACT: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoc@igf.org

Library are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: [http://www.tigr.org/tdb/bac\\_ends/mouse/Bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/Bac_end_intro.html)

Plate: 194 row: A column: 5

Seq primer: SP6

Class: BAC ends.

FEATURES

Location/Qualifiers

1..542

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-194A5"

/clone\_1db="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1;

ECORI; Site: 2; EORI: Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EORI and EORI methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the

ECORI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 149 a 97 c 84 g 212 t

alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AZ879168/rev ..

Align seg 1/1 to reverse of: AZ879168 from: 1 to: 542

41 GluLysGluAsnLysArgGluLeuLys 49  
|||||

522 GAGAAAGAGACAGAGAGAGCTAAAA 496

seq\_name: gb\_est2:BF338481

seq\_documentation\_block:

LOCUS BF338481

DEFINITION 602034145F2 NCI\_CGAP\_Brn64 Homo sapiens cDNA IMAGE:4182227

5', mRNA sequence.

ACCESSION BF338481

VERSION BF338481.1 GI:11284884

KEYWORDS EST.

SOURCE human.



ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 791)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-f@mail.nih.gov](mailto:cgapbs-f@mail.nih.gov)  
Tissue Procurement: David N. Louis, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM9496 row: 1 column: 12  
High quality sequence stop: 304.  
Location/Qualifiers  
1. 791  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI CGAP Brn64"  
/tissue\_type="gliblastoma with EGFR amplification"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: brain; Vector: PCMV-SPORT6; Site: 1; Note:  
Site: 2; Salt: Cloned unidirectionally; Primer: Oligo dT.  
Average insert size 1.57 kb. Constructed by Life  
Technologies. Note: this is a NCI CGAP Library."

BASE COUNT 412 a 104 c 160 g 115 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x BF338481 ..  
Align seg 1/1 to: BF338481 from: 1 to: 791

seq\_name: gb\_est2:BG533957  
52 LysylsGluysGluAspLysGlu 60  
|||||  
300 AAAAAAGAAAAGACAGAGAAAAGAA 326

seq\_documentation\_block:  
LOCUS BG533957 817 bp mRNA linear EST 03-APR-2001  
DEFINITION 60253056F1 NIH\_MGC\_77 Homo sapiens CDNA clone IMAGE:4662854 5',  
mRNA sequence.  
ACCESSION BG533957  
VERSION BG533957.1 GI:13525497  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 817)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-f@mail.nih.gov](mailto:cgapbs-f@mail.nih.gov)  
Tissue Procurement: Clontech Laboratories, Inc.  
CDNA Library Preparation: Clontech Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
Plate: LLMC1464 row: c column: 15  
High quality sequence stop: 730.  
Location/Qualifiers  
1. 817  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH-MGC\_77"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1;  
Site: 2; Salt: Cloned unidirectionally; Primer: Oligo dT.  
Average insert size 1.9  
kb (range 0.5-4.0 kb). 12/15 clones contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 139 c 193 g 233 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x BG533957 ..  
Align seg 1/1 to: BG533957 from: 1 to: 817

seq\_name: gb\_est2:BG612310  
63 GluLysAspAnLysGluLysArgGlu 71  
|||||  
251 GAAAAGCAATAGAGAAAAGAGAA 277

seq\_documentation\_block:  
LOCUS BG612310 817 bp mRNA linear EST 18-APR-2001  
DEFINITION 602613632F1 NIH\_MGC\_60 Homo sapiens CDNA clone IMAGE:4739191 5',  
mRNA sequence.  
ACCESSION BG612310  
VERSION BG612310.1 GI:13663681  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 817)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-f@mail.nih.gov](mailto:cgapbs-f@mail.nih.gov)  
Tissue Procurement: DCTP/DTP  
CDNA Library Preparation: Clontech Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLMC1609 row: p column: 08  
High quality sequence stop: 565.  
Location/Qualifiers  
1. 817  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH-MGC\_60"  
/tissue\_type="adenocarcinoma"

```

/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgctggcc); Site_2: SfII (ggccattagggc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCCGAGCGGCACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT      304 a      184 c      194 g      135 t
ORIGIN

alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x BG612310 ..

Align seg 1/1 to: BG612310 from: 1 to: 817

75 LysGluSerThrGlnLysGluLysGlu 83
|||||
657 AAGGAGAGCAGCAGCAAAAAGAGAAAGAG 683

seq_name: gb_est2:BF161157

seq_documentation_block:
LOCUS      BF161157          908 bp      mRNA      linear      EST 30-OCT-2000
DEFINITION  6012769402F1 NCL_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3968561 5',
mRNA sequence.
ACCESSION  BF161157
VERSION    BF161157.1  GI:11041264
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
1 (bases 1 to 908)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LRAM9197 row: d column: 02
High quality sequence stop: 620.
Location/Qualifiers
1. 908
/organism="Mus musculus"
/strain="Czech II (feral)"
/db_xref="taxon:10090"
/clone_1lb="IMAGE:3968561"
/clone_1lb="NCL_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI. Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

```

BASE COUNT      276 a      186 c      180 g      266 t
ORIGIN

alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x BF161157/rev ..

Align seg 1/1 to reverse of: BF161157 from: 1 to: 908

39 ThrLysGluLysGluAsnLysArgGlu 47
|||||
860 ACAAGGAGAGAGAAACAGAGAGAG 834

seq_name: gb_est2:BG435095

seq_documentation_block:
LOCUS      BG435095          1140 bp      mRNA      linear      EST 14-MAR-2001
DEFINITION  602507103F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4604160 5',
mRNA sequence.
ACCESSION  BG435095
VERSION    BG435095.1  GI:13341601
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1140)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Clontech Laboratories, Inc.
cDNA Library Preparation: Clontech Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LRCM1339 row: f column: 01
High quality sequence stop: 233.
Location/Qualifiers
1. 1140
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="IMAGE:4604160"
/clone_1lb="NIH_MGC_79"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgctggcc); Site_2: SfII (ggccattagggc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCCGAGCGGCACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT      520 a      206 c      341 g      73 t
ORIGIN

alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

```

## alignment\_block:

US-08-973-363-11 x BG435095

Align seg 1/1 to: BG435095 from: 1 to: 1149

30 LysylsGluAlaGluGluLysArgGlu 38

1061 AAGAAAGAGCGAGAGAGAGAGAGCA 1087

seq\_name: gb\_ests2:BG537228

## seq\_documentation\_block:

LOCUS BG537228 1149 bp mRNA linear EST 03-APR-2001

DEFINITION 602565257F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4689711 5',

RNA sequence.

ACCESSION BG537228

VERSION BG537228.1 GI:13529036

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1149)

AUTHORS NIH\_MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-rmail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LML at:

http://image.llnl.gov

Plate: L1CM1506 row: b column: 16

High quality sequence stop: 255.

Location/Qualifiers

## FEATURES

## source

1. 1149

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="NIH\_MGC\_77"

/lab\_host="DH10B (T1 phage-resistant)"

/note="organ: lung; vector: pBR322 (clontech); Site 1:

5' (ggcgccgcgcgcgc); Site 2: 5' (ggcgccgcgcgc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CAGGCGCATATGCGC-3' and 3' adaptor sequence:

5'-ATCTAGAGCGCGAGCGCGCAGCAGT-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH\_MGC Library."

BASE COUNT 504 a 187 c 400 g 58 t

## ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x BG537228

Align seg 1/1 to: BG537228 from: 1 to: 1149

32 GluAlaGluGluLysArgGluThrLys 40

459 GAGACGAGAGAGAGAGAGAGAGAGCA 485

seq\_name: gb\_gss:AG178168

## seq\_documentation\_block:

LOCUS AG178168 1310 bp DNA linear GSS 09-JAN-2002

DEFINITION Pan troglodytes DNA, clone: RP43-050A23.T7, genomic survey

sequence.

ACCESSION AG178168

VERSION AG178168.1 GI:16707848

KEYWORDS GSS; GSS (genome survey sequence).

SOURCE Pan troglodytes male lymphocytes DNA, clone: lib:RP43-Chimpanzee

Male BAC Library clone: RP43-050A23.T7.

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1 (sites)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Toto, Y., Watanabe, H., and Sakaki, Y.

BAC end sequences of Library RP43-43

Unpublished

2 (bases 1 to 1310)

TITLE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Toto, Y., Watanabe, H., and Sakaki, Y.

Submitted (02-APR-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC),

1-7-22, Suehiro-chou, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/)

Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library RP43-43 This BAC

end was generated during the Rad process and may have higher chance

of clone tracking errors.

PRIMERS

Sequencing: T7

LIBRARY

Vector : pBAC3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1. 1310

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone\_id="RP43-050A23.T7"

/sex="male"

/cell\_type="lymphocytes"

/clone\_id="RP43-Chimpanzee Male BAC Library"

BASE COUNT 37 a 448 c 371 g 436 t 18 others

## ORIGIN

## alignment\_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AG178168/rev

Align seg 1/1 to reverse of: AG178168 from: 1 to: 1310

30 LysylsGluAlaGluLysArgGlu 38

834 AAGAAAGAGCGAGAGAGAGAGAGCA 808

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 02:12:57 ; Search time 10310.5 seconds

(without alignments)  
537.850 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265  
Sequence: 1 GATGAGATTGTTTCACTGAA.....AAGAGTGAGAGAGAGAG 265

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenPdb1:\*  
1: gb\_ba:\*  
2: gb\_bkg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hggo\_inv:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

Result No.	Query Match	Length	DB ID	Description
1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	2 AF00733
4	26	9.8	192079	2 AC068141
5	26	9.8	237405	2 AC068142
6	24	9.1	52433	2 AC010197
7	24	9.1	101220	2 AC092372
8	24	9.1	134365	2 AC012624
9	24	9.1	143079	2 AC021449
10	24	9.1	145659	2 AC008531
11	24	9.1	15830	2 AC092736
12	24	9.1	154157	2 AF003509
13	24	9.1	157625	2 AC007194
14	24	9.1	193446	2 AC001946
15	24	9.1	219258	2 AC022121
16	24	9.1	276181	2 AC092382
17	23	8.7	137	6 A58694
18	23	8.7	9654	1 U67470
19	23	8.7	80374	8 T8K14
20	23	8.7	105680	2 AC006266
21	23	8.7	110184	2 AC020906
22	23	8.7	150874	2 AC016326
23	23	8.7	166396	2 AC023957
24	23	8.7	170885	2 AL356430
25	23	8.7	174135	2 AL359637
26	23	8.7	193101	2 AC009561
27	23	8.7	195837	2 AT0CHRIV21
28	23	8.7	198316	2 AC019157
29	23	8.7	203766	2 AC009592
30	22	8.3	1541	3 AF180393
31	22	8.3	1585	3 AF161311
32	22	8.3	12029	3 AE001432
33	22	8.3	12029	3 AE001432
34	22	8.3	14012	3 AE001433
35	22	8.3	83585	3 AL606475
36	22	8.3	83783	2 AC016513
37	22	8.3	92633	2 PFWAL4P1_3
38	22	8.3	104695	2 AC017784
39	22	8.3	110000	2 PFWAL4P1_0
40	22	8.3	114420	2 AC108194
41	22	8.3	117240	2 AC104479
42	22	8.3	121287	9 AC010332
43	22	8.3	130369	9 AC074090
44	22	8.3	135005	9 HS860F19
45	22	8.3	141504	2 AC022295

#### ALIGNMENTS

Result No.	Query Match	Length	DB ID	Description
1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	2 AF00733
4	26	9.8	192079	2 AC068141
5	26	9.8	237405	2 AC068142
6	24	9.1	52433	2 AC010197
7	24	9.1	101220	2 AC092372
8	24	9.1	134365	2 AC012624
9	24	9.1	143079	2 AC021449
10	24	9.1	145659	2 AC008531
11	24	9.1	15830	2 AC092736
12	24	9.1	154157	2 AF003509
13	24	9.1	157625	2 AC007194
14	24	9.1	193446	2 AC001946
15	24	9.1	219258	2 AC022121
16	24	9.1	276181	2 AC092382
17	23	8.7	137	6 A58694
18	23	8.7	9654	1 U67470
19	23	8.7	80374	8 T8K14
20	23	8.7	105680	2 AC006266
21	23	8.7	110184	2 AC020906
22	23	8.7	150874	2 AC016326
23	23	8.7	166396	2 AC023957
24	23	8.7	170885	2 AL356430
25	23	8.7	174135	2 AL359637
26	23	8.7	193101	2 AC009561
27	23	8.7	195837	2 AT0CHRIV21
28	23	8.7	198316	2 AC019157
29	23	8.7	203766	2 AC009592
30	22	8.3	1541	3 AF180393
31	22	8.3	1585	3 AF161311
32	22	8.3	12029	3 AE001432
33	22	8.3	12029	3 AE001432
34	22	8.3	14012	3 AE001433
35	22	8.3	83585	3 AL606475
36	22	8.3	83783	2 AC016513
37	22	8.3	92633	2 PFWAL4P1_3
38	22	8.3	104695	2 AC017784
39	22	8.3	110000	2 PFWAL4P1_0
40	22	8.3	114420	2 AC108194
41	22	8.3	117240	2 AC104479
42	22	8.3	121287	9 AC010332
43	22	8.3	130369	9 AC074090
44	22	8.3	135005	9 HS860F19
45	22	8.3	141504	2 AC022295

Result No.	Query Match	Length	DB ID	Description
1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	2 AF00733
4	26	9.8	192079	2 AC068141
5	26	9.8	237405	2 AC068142
6	24	9.1	52433	2 AC010197
7	24	9.1	101220	2 AC092372
8	24	9.1	134365	2 AC012624
9	24	9.1	143079	2 AC021449
10	24	9.1	145659	2 AC008531
11	24	9.1	15830	2 AC092736
12	24	9.1	154157	2 AF003509
13	24	9.1	157625	2 AC007194
14	24	9.1	193446	2 AC001946
15	24	9.1	219258	2 AC022121
16	24	9.1	276181	2 AC092382
17	23	8.7	137	6 A58694
18	23	8.7	9654	1 U67470
19	23	8.7	80374	8 T8K14
20	23	8.7	105680	2 AC006266
21	23	8.7	110184	2 AC020906
22	23	8.7	150874	2 AC016326
23	23	8.7	166396	2 AC023957
24	23	8.7	170885	2 AL356430
25	23	8.7	174135	2 AL359637
26	23	8.7	193101	2 AC009561
27	23	8.7	195837	2 AT0CHRIV21
28	23	8.7	198316	2 AC019157
29	23	8.7	203766	2 AC009592
30	22	8.3	1541	3 AF180393
31	22	8.3	1585	3 AF161311
32	22	8.3	12029	3 AE001432
33	22	8.3	12029	3 AE001432
34	22	8.3	14012	3 AE001433
35	22	8.3	83585	3 AL606475
36	22	8.3	83783	2 AC016513
37	22	8.3	92633	2 PFWAL4P1_3
38	22	8.3	104695	2 AC017784
39	22	8.3	110000	2 PFWAL4P1_0
40	22	8.3	114420	2 AC108194
41	22	8.3	117240	2 AC104479
42	22	8.3	121287	9 AC010332
43	22	8.3	130369	9 AC074090
44	22	8.3	135005	9 HS860F19
45	22	8.3	141504	2 AC022295

Result No.	Query Match	Length	DB ID	Description
1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	2 AF00733
4	26	9.8	192079	2 AC068141
5	26	9.8	237405	2 AC068142
6	24	9.1	52433	2 AC010197
7	24	9.1	101220	2 AC092372
8	24	9.1	134365	2 AC012624
9	24	9.1	143079	2 AC021449
10	24	9.1	145659	2 AC008531
11	24	9.1	15830	2 AC092736
12	24	9.1	154157	2 AF003509
13	24	9.1	157625	2 AC007194
14	24	9.1	193446	2 AC001946
15	24	9.1	219258	2 AC022121
16	24	9.1	276181	2 AC092382
17	23	8.7	137	6 A58694
18	23	8.7	9654	1 U67470
19	23	8.7	80374	8 T8K14
20	23	8.7	105680	2 AC006266
21	23	8.7	110184	2 AC020906
22	23	8.7	150874	2 AC016326
23	23	8.7	166396	2 AC023957
24	23	8.7	170885	2 AL356430
25	23	8.7	174135	2 AL359637
26	23	8.7	193101	2 AC009561
27	23	8.7	195837	2 AT0CHRIV21
28	23	8.7	198316	2 AC019157
29	23	8.7	203766	2 AC009592
30	22	8.3	1541	3 AF180393
31	22	8.3	1585	3 AF161311
32	22	8.3	12029	3 AE001432
33	22	8.3	12029	3 AE001432
34	22	8.3	14012	3 AE001433
35	22	8.3	83585	3 AL606475
36	22	8.3	83783	2 AC016513
37	22	8.3	92633	2 PFWAL4P1_3
38	22	8.3	104695	2 AC017784
39	22	8.3	110000	2 PFWAL4P1_0
40	22	8.3	114420	2 AC108194
41	22	8.3	117240	2 AC104479
42	22	8.3	121287	9 AC010332
43	22	8.3	130369	9 AC074090
44	22	8.3	135005	9 HS860F19
45	22	8.3	141504	2 AC022295

Result No.	Query Match	Length	DB ID	Description
1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	2 AF00733
4	26	9.8	192079	2 AC068141
5	26	9.8	237405	2 AC068142
6	24	9.1	52433	2 AC010197
7	24	9.1	101220	2 AC092372
8	24	9.1	134365	2 AC012624
9	24	9.1	143079	2 AC021449
10	24	9.1	145659	2 AC008531
11	24	9.1	15830	2 AC092736
12	24	9.1	154157	2 AF003509
13	24	9.1	157625	2 AC007194
14	24	9.1	193446	2 AC001946
15	24	9.1	219258	2 AC022121
16	24	9.1	276181	2 AC092382
17	23	8.7	137	6 A58694
18	23	8.7	9654	1 U67470
19	23	8.7	80374	8 T8K14
20	23	8.7	105680	2 AC006266
21	23	8.7	110184	2 AC020906
22	23	8.7	150874	2 AC016326
23	23	8.7	166396	2 AC023957
24	23	8.7	170885	2 AL356430
25	23	8.7	174135	2 AL359637
26	23	8.7	193101	2 AC009561
27	23	8.7	195837	2 AT0CHRIV21
28	23	8.7	198316	2 AC019157
29	23	8.7	203766	2 AC009592
30	22	8.3	1541	3 AF180393
31	22	8.3	1585	3 AF161311
32	22	8.3	12029	3 AE001432
33	22	8.3	12029	3 AE001432
34	22	8.3	14012	3 AE001433
35	22	8.3	83585	3 AL606475
36	22	8.3	83783	2

Query Match 100.0%; Score 265; DB 6; Length 265;  
Best Local Similarity 100.0%; Pred. No. 1.3e-106;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGAGATTGTTTCAGTGAAACATCTACATTAATAAATAAAGAGAAAAAATGA 60  
DB 1 GATGAGATTGTTTCAGTGAAACATCTACATTAATAAATAAAGAGAAAAAATGA 60  
QY 61 AGAAACCGTGGACCATATGTTATTAAGAGAGCTAAGAAAAAGAGACAAA 120  
DB 61 AGAAACCGTGGACCATATGTTATTAAGAGAGCTAAGAAAAAGAGACAAA 120  
QY 121 AGAGAGAAATTAAGAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 121 AGAGAGAGAAATTAAGAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 ATTAAG 240  
DB 181 ATTAAG 240  
QY 241 AG 265  
DB 241 AG 265

## RESULT 2

AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
LOCUS Gallus gallus chemo-helicase-DNA-binding on the Z chromosome  
DEFINITION protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cDS.

ACCESSION AF004397  
VERSION AF004397.1 GI:2501845

## KEYWORDS

## SOURCE

ORGANISM

chicken.  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

## REFERENCE

1 (bases 1 to 6872)  
Griffiths, R. and Korn, R.M.  
A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
Gene 197 (1-2), 225-229 (1997)

## JOURNAL

## MEDLINE

## AUTHORS

## TITLE

## REFERENCE

## JOURNAL

## FEATURES

## source

gene  
CDS

1.6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1.6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chemo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501845"  
/translation="MNGHSDSESVRNSGSSSRDDSGSAGSGSGSSSGSSDGGSS  
SOGSSDSESGSSGSSSESDTSREKQVOKPKPADSEFKSSPSIIAYVRSAY  
LKKQOQKAAASDSSGSEEDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS  
SSSAAEDGKASCESESDYEPKNNKVRKRPSPKIRKSGKSTGOKKROLDSEEE  
DDDEDYKGRSROATYVNSKAEKRTKSDILKPCGSDYVPTDEDEFTEIKFMD  
SRIGRKATGASTTITAVEADGDPNAGFEKSEKELGELYIILKMGSHINTWETET  
LKQONVGMNKLNDYKKKDDKTRKMLKNASPEDEVYNNCOELTDLHKQYVIERII

## misc\_feature

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

Query Match 82.3%; Score 218; DB 5; Length 6872;  
Best Local Similarity 100.0%; Pred. No. 4.8e-86;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 AAAAAAATGAAAGAAAGCCCTGACCATGTTGTTATTAAGAGAGAGCTGAGAAA 107  
DB 4387 AAAAAAATGAAAGAAAGCCCTGACCATGTTGTTATTAAGAGAGAGCTGAGAAA 4446  
QY 108 AAG 167  
DB 4447 AAG 4506  
QY 168 AGCATTAAGAAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227  
DB 4507 AGCATTAAGAAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4566  
QY 228 AATCCACACAG 265  
DB 4567 AATCCACACAG 4604

## RESULT 3

## AC100733

## LOCUS

## DEFINITION

## UNORDERED

## PIECES.

## AC100733

## AC100733.1

## HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

## KEYWORDS

## SOURCE

## ORGANISM

154919 bp DNA linear HTG 22-NOV-2001  
Mus musculus clone RP24-312H21, WORKING DRAFT SEQUENCE, 17  
unpublished pieces.  
AC100733  
AC100733.1 GI:17048100  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.

1 (bases 1 to 154919)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
JOURNAL  
TITLE  
Unpublished  
2 (bases 1 to 154919)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Batra, N., Bastien, V., Boguski, L., Boukigal, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

TITLE  
JOURNAL  
COMMENT

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gargyala, S.,  
Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Labouque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliwer, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V.,  
Raymond, C., Retter, R., Rieback, M., Riley, R., Riss, C., Rogov, E.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seman, S., Severi, F., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Straus, N., Sudriamanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L17445  
Center clone name: 312\_H-21

Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 149301 bases at least Q40  
Consensus quality: 151798 bases at least Q30  
Consensus quality: 152705 bases at least Q20  
Insert size: 151000; agarose-fp  
Insert size: 153319; sum-of-ctrls  
Quality coverage: 7.0 in Q20 bases; agarose-fp  
Quality coverage: 6.8 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently  
consists of 17 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 23835: contig of 23835 bp in length  
\* 23836 23935: gap of 100 bp  
\* 23936 25171: contig of 1236 bp in length  
\* 25172 25271: gap of 100 bp  
\* 25272 26348: contig of 1077 bp in length  
\* 26349 26448: gap of 100 bp  
\* 26449 29041: contig of 2593 bp in length  
\* 29042 29141: gap of 100 bp  
\* 29142 31763: contig of 2622 bp in length  
\* 31764 31863: gap of 100 bp  
\* 31864 33051: contig of 1188 bp in length  
\* 33052 33151: gap of 100 bp  
\* 33152 35512: contig of 2361 bp in length  
\* 35513 35612: gap of 100 bp  
\* 35613 37576: contig of 1964 bp in length  
\* 37577 37676: gap of 100 bp  
\* 37677 40717: contig of 3041 bp in length  
\* 40718 40817: gap of 100 bp  
\* 40818 43984: contig of 3167 bp in length  
\* 43985 44084: gap of 100 bp

FEATURES  
source  
\* 44085 48175: contig of 4091 bp in length  
\* 48176 48275: gap of 100 bp  
\* 48276 53706: contig of 5431 bp in length  
\* 53707 53806: gap of 100 bp  
\* 53807 59511: contig of 5705 bp in length  
\* 59512 59611: gap of 100 bp  
\* 59612 72539: contig of 12928 bp in length  
\* 72540 72639: gap of 100 bp  
\* 72640 86593: contig of 13954 bp in length  
\* 86594 86693: gap of 100 bp  
\* 86694 111426: contig of 24733 bp in length  
\* 111427 111526: gap of 100 bp  
\* 111527 1154919: contig of 43393 bp in length.  
Location/Qualifiers  
1. 154919  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP24-312H21"  
/clone="lib-RPCT-24 Male Mouse BAC"  
1. 23835  
/note="assembly-fragment"  
clone\_end:SP6  
vector\_side:left  
23936. 25171  
/note="assembly-fragment"  
25272. 26348  
/note="assembly-fragment"  
26449. 29041  
/note="assembly-fragment"  
29142. 31763  
/note="assembly-fragment"  
31864. 33051  
/note="assembly-fragment"  
33152. 35512  
/note="assembly-fragment"  
35613. 37576  
/note="assembly-fragment"  
37677. 40717  
/note="assembly-fragment"  
40818. 43984  
/note="assembly-fragment"  
44085. 48175  
/note="assembly-fragment"  
/note="assembly-fragment"  
48276. 53706  
/note="assembly-fragment"  
53807. 59511  
/note="assembly-fragment"  
59612. 72539  
/note="assembly-fragment"  
72640. 86593  
/note="assembly-fragment"  
86694. 111426  
/note="assembly-fragment"  
111527. 1154919  
/note="assembly-fragment"  
clone\_end:77  
vector\_side:right

BASE COUNT 49030 a 29885 c 29043 g 45353 t 1608 others

ORIGIN

Query Match 9.8%; Score 26; DB 2; Length 154919;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AAAAAATATAAAAACGAAAAACGAAA 56  
|||||  
DB 11250 AAAAAATATAAAAACGAAAAACGAAA 11275

RESULT 4  
AC068141 192079 bp DNA linear HTG 29-APR-2000  
LOCUS AC068141

```
DEFINITION Mus musculus chromosome 5 clone RP23-326E9 strain C57BL6/J, WORKING
ACCESSION AC068141
VERSION AC068141.1 GI:7670125
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192079)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantiripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 192079)
Green,E.D.
Direct Submission
Submitted (29-APR-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
Project Information
Center clone name: 326E09
Center project name: yu
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187247 bases at least Q40
Consensus quality: 188227 bases at least Q30
Consensus quality: 188721 bases at least Q20
Insert size: 177000; agarose-gel
Insert size: 186000; pulse-field-gel
Insert size: 192079; sum-of-ctrls
Quality coverage: 8.09x in Q20 bases; agarose-gel
Quality coverage: 7.69x in Q20 bases; pulse-field-gel
Quality coverage: 7.45x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6760: contig of 6760 bp in length
6761 20444: gap of unknown length
6761 20444: contig of 13684 bp in length
20445 36459: gap of unknown length
20445 36459: contig of 16015 bp in length
36460 58509: gap of unknown length
36460 58509: contig of 22050 bp in length
58510 81567: gap of unknown length
58510 81567: contig of 23058 bp in length
81568 115183: gap of unknown length
81568 115183: contig of 33616 bp in length
115184 149813: gap of unknown length
115184 149813: contig of 34630 bp in length
149814 192079: gap of unknown length
149814 192079: contig of 42266 bp in length.
location/Qualifiers
1. 192079
/organism="Mus musculus"
/strain="C57BL6/J"

/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-326E9"
/clone_id="RPCT mouse BAC library 23"
BASE COUNT 58560 a 36452 c 35882 g 61161 t 24 others
ORIGIN
Query Match 9.8%; Score 26; DB 2; Length 192079;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;
QY 32 AAAAAATAAAAACAGAAAAAGAAAA 57
Db 121716 AAAAAATAAAAACAGAAAAAGAAAA 121741

RESULT 5
AC068142/c
LOCUS 237405 bp DNA linear HTG 29-APR-2000
DEFINITION Mus musculus chromosome 5 clone RP23-53F18 strain C57BL6/J, WORKING
DRAFT SEQUENCE, 9 unordered pieces.
AC068142
VERSION AC068142.1 GI:7670126
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 237405)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantiripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 237405)
Green,E.D.
Direct Submission
Submitted (29-APR-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
Project Information
Center clone name: yu
Center project name: yt
Center clone name: 053F18
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 230333 bases at least Q40
Consensus quality: 232224 bases at least Q30
Consensus quality: 233239 bases at least Q20
Insert size: 212000; agarose-gel
Insert size: 249000; pulse-field-gel
Insert size: 237405; sum-of-ctrls
Quality coverage: 7.28x in Q20 bases; agarose-gel
Quality coverage: 6.20x in Q20 bases; pulse-field-gel
Quality coverage: 6.50x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
```



```

* be preserved.
* 1 4718: contig of 4718 bp in length
* 4719 11420: contig of 6702 bp in length
* 11421 28006: contig of 16586 bp in length
* 28007 48211: contig of 20205 bp in length
* 48212 75969: contig of 27758 bp in length
* 75970 110435: contig of 34466 bp in length
* 110436 145778: contig of 35343 bp in length
* 145779 180307: contig of 34529 bp in length
* 180308 237405: contig of 57098 bp in length.
Location/Qualifiers
1. 237405
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-53F18"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 72271 a 44891 c 45742 g 74492 t 9 others
ORIGIN

```

```

Query Match 9.8%; Score 26; DB 2; Length 237405;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 32 AAAAAATTAACGAAAAAGAAA 57
DB 173891 AAAAAATTAACGAAAAAGAAA 173866

```

```

RESULT 6
AC101977 52433 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP24-333M7, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC101977.1 GI:17060753
VERSION AC101977.1 GI:17060753
KEYWORDS HTG, HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-333M7
Unpublished
2 (bases 1 to 52433)

```

```

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepe,L., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Garaya,S.,
Ginde,S., Gord,S., Goyette,M., Graham,W., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,
Lamaszars,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Meijer,J., Marcus,N., Matthews,C.,
McCarthy,M., McKean,P., McKenna,K., McSheeters,R., Meldrum,J.,
Menon,L., Mihova,T., Minge,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliyer,J., Peterson,K., Phunhahng,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

```

# TITLE JOURNAL

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project information

Center project name: 117818

Center clone name: 333M\_7

NOTE: This record contains 67 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

```

1 754: contig of 754 bp in length
755 854: gap of 100 bp
855 1486: contig of 632 bp in length
1487 1586: gap of 100 bp
1587 2163: contig of 577 bp in length
2164 2263: gap of 100 bp
2264 2968: contig of 705 bp in length
2969 3068: gap of 100 bp
3069 3774: contig of 706 bp in length
3775 3874: gap of 100 bp
3875 4524: contig of 650 bp in length
4525 4624: gap of 100 bp
4625 5301: contig of 677 bp in length
5302 5401: gap of 100 bp
5402 6084: contig of 683 bp in length
6085 6184: gap of 100 bp
6185 6864: contig of 680 bp in length
6865 6964: gap of 100 bp
6965 7634: contig of 670 bp in length
7635 7734: gap of 100 bp
7735 8437: contig of 703 bp in length
8438 8537: gap of 100 bp
8538 9236: contig of 699 bp in length
9237 9336: gap of 100 bp
9337 9994: contig of 658 bp in length
9995 10094: gap of 100 bp
10095 10794: contig of 700 bp in length
10795 10894: gap of 100 bp
10895 11583: contig of 689 bp in length
11584 11683: gap of 100 bp
11684 12362: contig of 679 bp in length
12363 12462: gap of 100 bp
12463 13145: contig of 683 bp in length
13146 13245: gap of 100 bp
13246 13931: contig of 686 bp in length
13932 14031: gap of 100 bp
14032 141703: contig of 672 bp in length
14704 14803: gap of 100 bp
14804 15467: contig of 664 bp in length
14805 15567: gap of 100 bp
15568 16269: contig of 702 bp in length
16270 16369: gap of 100 bp
16370 17042: contig of 673 bp in length

```

\* 17043 17142: gap of 100 bp  
\* 17143 17845: contig of 703 bp in length  
\* 17846 17945: gap of 100 bp  
\* 17946 18708: contig of 764 bp in length  
\* 18710 18809: gap of 100 bp  
\* 18810 19500: contig of 691 bp in length  
\* 19501 19600: gap of 100 bp  
\* 19601 20277: contig of 677 bp in length  
\* 20278 20377: gap of 100 bp  
\* 20378 21054: contig of 677 bp in length  
\* 21055 21154: gap of 100 bp  
\* 21155 21831: contig of 677 bp in length  
\* 21832 21931: gap of 100 bp  
\* 21932 22622: contig of 691 bp in length  
\* 22623 22723: gap of 100 bp  
\* 22723 23398: contig of 676 bp in length  
\* 23399 23498: gap of 100 bp  
\* 23499 24212: contig of 714 bp in length  
\* 24213 24312: gap of 100 bp  
\* 24313 24975: contig of 663 bp in length  
\* 24976 25075: gap of 100 bp  
\* 25076 25800: contig of 725 bp in length  
\* 25801 25900: gap of 100 bp  
\* 25901 26600: contig of 700 bp in length  
\* 26601 26700: gap of 100 bp  
\* 26701 27379: contig of 679 bp in length  
\* 27380 27479: gap of 100 bp  
\* 27480 28151: contig of 672 bp in length  
\* 28152 28251: gap of 100 bp  
\* 28252 28840: contig of 589 bp in length  
\* 28841 28940: gap of 100 bp  
\* 28941 29622: contig of 682 bp in length  
\* 29623 29722: gap of 100 bp  
\* 29723 30383: contig of 661 bp in length  
\* 30384 30483: gap of 100 bp  
\* 30484 31152: contig of 669 bp in length  
\* 31153 31252: gap of 100 bp  
\* 31253 31951: contig of 699 bp in length  
\* 31952 32051: gap of 100 bp  
\* 32052 32746: contig of 695 bp in length  
\* 32747 32846: gap of 100 bp  
\* 32847 33543: contig of 703 bp in length  
\* 33550 33649: gap of 100 bp  
\* 33650 34355: contig of 706 bp in length  
\* 34356 34455: gap of 100 bp  
\* 34456 35169: contig of 714 bp in length  
\* 35170 35269: gap of 100 bp  
\* 35270 35922: contig of 653 bp in length  
\* 35923 36022: gap of 100 bp  
\* 36023 36707: contig of 685 bp in length  
\* 36708 36807: gap of 100 bp  
\* 36808 37491: contig of 684 bp in length  
\* 37492 37591: gap of 100 bp  
\* 37592 38263: contig of 672 bp in length  
\* 38264 38363: gap of 100 bp  
\* 38364 39073: contig of 710 bp in length  
\* 39074 39173: gap of 100 bp  
\* 39174 39832: contig of 659 bp in length  
\* 39833 39932: gap of 100 bp  
\* 39933 40643: contig of 711 bp in length  
\* 40644 40743: gap of 100 bp  
\* 40744 41441: contig of 698 bp in length  
\* 41442 41541: gap of 100 bp  
\* 41542 42200: contig of 659 bp in length  
\* 42201 42300: gap of 100 bp  
\* 42301 42977: contig of 677 bp in length  
\* 42978 43077: gap of 100 bp  
\* 43078 43760: contig of 683 bp in length  
\* 43761 43860: gap of 100 bp  
\* 43861 44557: contig of 697 bp in length  
\* 44558 44657: gap of 100 bp  
\* 44658 45283: contig of 626 bp in length  
\* 45284 45383: gap of 100 bp

## FEATURES

source

\* 45384 46089: contig of 706 bp in length  
\* 46090 46189: gap of 100 bp  
\* 46190 46860: contig of 671 bp in length  
\* 46861 46960: gap of 100 bp  
\* 46961 47688: contig of 728 bp in length  
\* 47689 47788: gap of 100 bp  
\* 47789 48467: contig of 679 bp in length  
\* 48468 48567: gap of 100 bp  
\* 48568 49254: contig of 687 bp in length  
\* 49255 49354: gap of 100 bp  
\* 49355 50021: contig of 667 bp in length  
\* 50022 50121: gap of 100 bp  
\* 50122 50825: contig of 704 bp in length  
\* 50826 50925: gap of 100 bp  
\* 50926 51628: contig of 703 bp in length  
\* 51629 51728: gap of 100 bp  
\* 51729 52433: contig of 705 bp in length.  
Location/Qualifiers  
1..52433  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP24-333M7"

Query Match 9.1%; Score 24; DB 2; Length 52433;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 ATAAAAAATATAAACGACAAAAA 52

Db 7863 ATAAAAAATATAAACGACAAAAA 7886

## RESULT 7

AC092372/c 101220 bp DNA linear PRI 07-DEC-2001  
LOCUS Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
DEFINITION AC092372

AC092372.3 GI:17402768

VERSION HTG.

KEYWORDS human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 101220)

REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Unpublished

2 (bases 1 to 101220)  
DOE Joint Genome Institute.

REFERENCE Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 101220)

DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA  
On Dec 7, 2001 this sequence version replaced gi:15290448.

COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.

NOTE: This insert is not the entire sequence of the clone (entire  
sequence is 146,7kd). It is clipped at the overlap with AC012624.

The number of bases overlapped is 90404.  
Location/Qualifiers  
1..101220

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/chromosome="5"  
/clone="RP11-58M12"

## FEATURES

source

BASE COUNT	34122 a	18862 c	17827 g	30409 f
ORIGIN				

\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc-feature  
1. 38820  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
38921. 40411  
/note="assembly\_fragment"  
40512. 43279  
/note="assembly\_fragment"  
43380. 46905  
/note="assembly\_fragment"  
47006. 51830  
/note="assembly\_fragment"  
51931. 62619  
/note="assembly\_fragment"  
62720. 75408  
/note="assembly\_fragment"  
75509. 92516  
/note="assembly\_fragment"  
92617. 106409  
/note="assembly\_fragment"  
106510. 143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right  
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 9.1%; Score 24; DB 2; Length 143079;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 AAAAGAAAAAGAGATAGAGAGA 180  
|||||  
Db 117824 AAAAGAAAAAGAGATAGAGAGA 117847

RESULT 10  
AC008531 145659 bp DNA linear HTG 14-FEB-2001  
LOCUS Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
DEFINITION 7 ordered pieces.  
ACCESSION AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 145659)  
JOURNAL DOE Joint Genome Institute.  
REFERENCE Direct Submission  
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI

Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 56174: contig of 56174 bp in length  
\* 56175 56274: gap of unknown length  
\* 56275 100874: contig of 44600 bp in length  
\* 100875 100974: gap of unknown length  
\* 100975 113127: contig of 12153 bp in length  
\* 113128 113227: gap of unknown length  
\* 113228 118190: contig of 4963 bp in length  
\* 118191 118290: gap of unknown length  
\* 118291 119694: contig of 1404 bp in length  
\* 119695 119794: gap of unknown length  
\* 119795 123297: contig of 3503 bp in length  
\* 123298 123398: gap of unknown length  
\* 123398 145659: contig of 22262 bp in length.  
Location/Qualifiers  
1. 145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-480B11"  
/clone\_lib="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 9.1%; Score 24; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 AAAAGAAAAAGAGATAGAGAGA 180  
|||||  
Db 48441 AAAAGAAAAAGAGATAGAGAGA 48464

RESULT 11  
AC092736/c 145830 bp DNA linear HTG 04-OCT-2001  
LOCUS Trypanosoma brucei chromosome IX clone RPC193-28616, \*\*\* SEQUENCING  
DEFINITION IN PROGRESS \*\*\*, 1 ordered pieces.  
ACCESSION AC092736  
VERSION AC092736.5 GI:15920106  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 145830)  
REFERENCE El-Sayed, N.M., Gheddi, E., Song, J., Larkin, C., Wanless, D., Jones, K.,  
JOURNAL Peterson, J., Hou, L., Zhao, H., Mason, T., Mollis, J., Pal, G., Van  
Aken, S., Ullrich, T., Khalak, H.G., Gerard, C., Leech, V., Dillu, E.,  
Melville, S., White, O., Adams, M.D., Donelson, J.E. and Fraser, C.M.

TITLE Trypanosoma brucei GUT10.1 RPI193-28616 BAC genomic sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145830)  
AUTHORS El-Sayed, N.M., Khalak, H. and Adams, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-2001) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Oct 4, 2001 this sequence version replaced g1:15789236.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source 1 145830 contig of 145830 bp in length.  
1 145830  
Location/Qualifiers  
1 145830  
/organism="Trypanosoma brucei"  
/isolate="GUT10.1"  
/db\_xref="taxon:5691"  
/chromosome="IX"  
/clone="RPI193-28616"  
BASE COUNT 35876 a 32464 c 37241 g 40249 t  
ORIGIN  
Query Match 9.1%; Score 24; DB 2; Length 145830;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 172 TAAGAAAGATTAAGAAAGAAAGA 195  
|||||  
Db 139024 TAAGAAAGATTAAGAAAGAAAGA 139001  
RESULT 12  
AP003509 154157 bp DNA linear HTG 12-APR-2001  
LOCUS Oryza sativa chromosome 6 clone P0525F01, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AP003509  
VERSION AP003509.1 GI:13603469  
KEYWORDS HTG: HTGS PHASE2.  
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0525F01.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehharitidae; Oryzae; Oryza.  
1 (slices)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa Nipponbare (GA3) genomic DNA, chromosome 6, PAC  
clone:P0525F01  
2 (bases 1 to 154157)  
Published Only in Database (2001) In press  
JOURNAL Direct Submission  
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
TITLE Submitted (11-APR-2001) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
FEATURES  
source 1 154157  
Location/Qualifiers  
1 154157  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="6"  
/clone="P0525F01"  
BASE COUNT 43756 a 32256 c 33371 g 44624 t 150 others  
ORIGIN  
Query Match 9.1%; Score 24; DB 2; Length 154157;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 ATTAAGAAATTAAGAAAGAAAGAA 52  
|||||  
Db 35211 ATTAAGAAATTAAGAAAGAAAGAA 35234  
RESULT 13  
AC007934 157625 bp DNA linear PRI 21-JAN-2000  
LOCUS Homo sapiens, clone RP11-29A1, complete sequence.  
DEFINITION AC007934  
ACCESSION AC007934  
VERSION AC007934.7 GI:6721208  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 157625)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Unpublished  
2 (bases 1 to 157625)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Baran, N., Beckert, R., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collumore, A.,  
Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,  
Domelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gaidyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heath, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Karatas, A., Lebecky, J., Lien, C., Locke, K., Macdonald, P.,  
Marquis, N., McKean, P., McKen, A., McKernan, K., McLaughlin, J.,  
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
Naylor, J., Nijhoff, M., O'Connor, T., O'Donnell, P., Pavlid, B.,  
Peterson, K., Pollard, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Testa, S., Toriella Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (30-JUN-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 157625)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Baran, N., Beckert, R., Benn, J., Brown, A.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,  
Chopel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domelan, L., Doyle, M., Feneator, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gaidyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, D.,  
Lander, E., Lebecky, J., Levine, R., Lien, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McKean, P., McKen, A., McKernan, K.,  
McNeeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisan, C., Pollard, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (21-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Jan 20, 2000 this sequence version replaced gi:5532078.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L601  
Center clone name: 29\_A1

```
FEATURES
Source          Location/Qualifiers
1..157625
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPC1-11 Human Male BAC"
/repeat_region complement(79..165)
/rpt_family="MER41B"
166..466
/rpt_family="Alusg"
/repeat_region complement(467..907)
/rpt_family="MER41B"
/repeat_region complement(933..1099)
/rpt_family="MER41D"
/repeat_region complement(1100..1209)
/rpt_family="HERVK14|HERVK1"
/repeat_region complement(1215..1469)
/rpt_family="MER66-internal"
/repeat_region complement(1820..2044)
/rpt_family="MER66-internal"
/repeat_region complement(2048..2153)
/rpt_family="MER49"
/repeat_region complement(2139..2426)
/rpt_family="MER49"
/repeat_region complement(2550..2691)
/rpt_family="MER66-internal"
/repeat_region complement(2819..2908)
/rpt_family="MER4-internal"
/repeat_region complement(3119..3145)
/rpt_family="MER4-internal"
3146..3453
/rpt_family="Alusx"
/repeat_region complement(3454..3612)
/rpt_family="MER4-internal"
/repeat_region complement(3797..3951)
/rpt_family="MER4-internal"
/repeat_region complement(4004..4094)
/rpt_family="Alusg"
/repeat_region complement(4095..4875)
/rpt_family="LTR17"
/repeat_region complement(4876..5084)
/rpt_family="Alusg"
/repeat_region complement(5102..5311)
/rpt_family="MER4-internal"
/repeat_region complement(5146..5362)
/rpt_family="MER83-internal"
5363..5669
/rpt_family="Aluy"
/repeat_region complement(5670..5954)
/rpt_family="MER83-internal"
5962..6259
/rpt_family="Alusg"
/repeat_region complement(6264..6564)
/rpt_family="Alusg"
6565..6598
/rpt_family="AT-rich"
6599..6777
/rpt_family="L1"
/repeat_region complement(6782..6914)
/rpt_family="MER83-internal"
6913..6969
/rpt_family="MER51-internal"
complement(6970..7269)
/rpt_family="Alusg"
7270..7998
/rpt_family="MER51-internal"
7999..8019
/rpt_family="(TG)n"
8020..8387
/rpt_family="MER51-internal"
8588..8886
/rpt_family="Alusx"
8887..9231
/rpt_family="MER51-internal"
9232..10683
/rpt_family="L1Pa6"
10691..11132
/rpt_family="L1Pa6"
11147..12189
/rpt_family="MER51-internal"
12215..12923
/rpt_family="MER57-internal"
complement(12932..13235)
/rpt_family="Alusg"
13242..13482
/rpt_family="MER51B"
13531..13837
/rpt_family="Aluy"
13843..14226
/rpt_family="MER51A"
complement(14280..14992)
/rpt_family="LTRR8A"
15702..15750
/rpt_family="(T)n"
complement(15791..15820)
/rpt_family="MIR"
complement(15943..16110)
/rpt_family="L1Pa4"
16257..16285
/rpt_family="AT-rich"
complement(17072..17391)
/rpt_family="AlusC"
17758..17806
/rpt_family="CT-rich"
17873..17903
/rpt_family="(CA)n"
18042..18265
/rpt_family="MER112"
complement(19327..19729)
/rpt_family="LTR1J"
complement(19733..20084)
/rpt_family="THEIC"
20085..20375
/rpt_family="Aluy"
complement(20392..20480)
/rpt_family="LTR1J"
21388..21554
/rpt_family="MIR"
21509..21564
/rpt_family="L2"
21942..21984
/rpt_family="AT-rich"
22283..22941
/rpt_family="L1"
24397..24591
/rpt_family="L1"
complement(24594..24971)
/rpt_family="LTR1C"
24980..25092
/rpt_family="L1"
```

```

repeat_region      25106..25949
                   /rpt_family="t1m1"
repeat_region      25955..26103
                   /rpt_family="t1"

Query Match      9 18; Score 24; DB 9; Length 157625;
Best Local Similarity 100.0%; Freq. NO 111;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0

33 AAAAATTAACGAGAAAAGCAA 56
|||||
Db 26698 AAAAATTAACGAGAAAAGCAA 28721

RESULT 14
AC091946
LOCUS      AC091946      193446 bp      DNA      linear      HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012. *** SEQUENCING IN
AC091946
AC091946      AC091946      33 unordered pieces.
AC091946.1 GI:14333882
HTG: HTGS_PHASE1.
human.
Organism      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 193446)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 193446)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
--Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web Site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center Clone name: RPC1-11_36012
-----
Summary Statistics
Consensus quality: 15767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1358: contig of 1358 bp in length
*      1359      1458: gap of unknown length
*      1459      2667: contig of 1209 bp in length
*      2668      2767: gap of unknown length
*      2768      3823: contig of 1055 bp in length
*      3824      3923: gap of unknown length
*      3924      5920: contig of 1997 bp in length
*      5921      6021: gap of unknown length
*      6021      7494: contig of 1474 bp in length
*      7495      7595: gap of unknown length
*      7596      9097: contig of 1502 bp in length
*      9097      10260: gap of unknown length
*      10260: contig of 1064 bp in length

```

Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
157	100.0%	24	9.18	89335	193446	0	0
<pre> 157 AAAAAAAAAAGGATTAAGAAAGA 180       Db 89335 AAAAAAAAAAGGATTAAGAAAGA 89335 </pre>							

AC022121/c  
 LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
 AC022121  
 AC022121.6 GI:15375145  
 HG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS 1 (bases 1 to 219258)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission  
 REFERENCE 2 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission

REFERENCE  
 AUTHORS Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
 TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 JOURNAL 3 (bases 1 to 219258)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission

REFERENCE  
 AUTHORS Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 TITLE Drive, Walnut Creek, CA 94598, USA  
 JOURNAL 4 (bases 1 to 219258)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission

REFERENCE  
 AUTHORS Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 TITLE Drive, Walnut Creek, CA 94598, USA  
 JOURNAL On Aug 30, 2001 this sequence version replaced gi:15148108.  
 COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of Sequence;  
 Estimated Total Number of Errors is 0.4.  
 STS Content:  
 WI-5811 G04974  
 WI-13675 G23101  
 SHGC-58345 G38487  
 SHGC-103595 G57841.

FEATURES  
 Source Location/Qualifiers  
 1. .219258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2007H13"  
 BASE COUNT 71954 a 42062 c 40933 g 64309 t  
 ORIGIN

Query Match 9.1%; Score 24; DB 9; Length 219258;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 157 AAAAGAAAAAGAGATAAGAAAGA 180  
 ||||||||||||||||||  
 Db 87769 AAAAGAAAAAGAGATAAGAAAGA 87746

Search completed: August 3, 2002, 02:15:42  
 Job time: 42712 sec





PI Griffiths R, Tiwari B;  
 XX WPI: 1997-043127/04.  
 DR P-PSDB; AAW08144.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Disclosure; Fig 7; 76pp; English.  
 XX  
 CC A composite nucleotide sequence (AAW42752) and putative translation  
 CC (AAW08144) sequence are provided of a motif that is found spliced to  
 CC a proportion of chicken CHD-1A clones. The motif is inserted  
 CC between bases 4327 and 4328 of the CHD-1A composite sequence  
 CC (AAW42751). None of the 7 CHD-1 clones examined contained the  
 CC complete motif. There are no splice donor or acceptor sites within  
 CC the motif suggesting it is a final rather than an intermediary  
 CC product of splicing. The motif is also found at the 3' end of the  
 CC CHD-W clone CC14 (see also AAT42753).  
 CC  
 XX Sequence 265 BP; 158 A; 16 C; 61 G; 30 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 265; DB 18; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-101;  
 Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GATGAGATTGTTTCAGTGAACATCTACATTAATAAATGAAGAAAAATGA 60  
 Db 1 gatgagattgtttcagtgaaacatctacataaataaagaaagaaatga 60  
 QY 61 AGAAAGCCTGAGCCAGATATTGTTATPAAGAGAAAGCTGAGAGAAAAAGAGACAA 120  
 Db 61 agaaagcctgagccagatattgtgtataaagaagagctgaagaaagaaagagacaa 120  
 QY 121 AGGAGAGAGAAATTAAGGGAATTGAAGAGAGAAAAAAGAGATTAAGAAAGA 180  
 Db 121 aggaagagaaataaaggaattgaaaggaagaaagaaagagagataagaaga 180  
 QY 181 ATTAAAGAAAAAGTAAATTAAGAAAGAGAGAAACAAATTAAGATCCACAGAGAA 240  
 Db 181 attaaagaaaaagatatataaagagagagaaacaaagtaaaagatccacacagaa 240  
 QY 241 AGAAAGAAAGTGAAGAGAGAGAG 265  
 Db 241 agaaagaaagtgagagagagagag 265  
 RESULT 2  
 AAT42753  
 ID AAT42753 standard; cDNA; 137 BP.  
 AC AAT42753;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-W clone CC14 3' motif.  
 XX  
 KW Bird; sex determination; chromodomain-helicase-DNA binding 1 Avian;  
 KW CHD-W; chromodomain-helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key location/Qualifiers  
 FT misc\_difference 52  
 FT /\*tag- a  
 FT /note= "base 52 disrupts the reading frame for  
 FT the translated amino acid sequence given  
 FT in Fig 7"  
 XX  
 XX WC09639505-A1.  
 XX  
 PD 12-DEC-1996.

XX  
 PF 05-JUN-1996; 96WO-GB01341.  
 XX  
 XX 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 XX WPI: 1997-043127/04.  
 DR P-PSDB; AAW08145.  
 XX  
 CC Avian chromodomain-helicase-DNA binding genes determine sex in  
 CC birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Disclosure; Fig 7; 76pp; English.  
 XX  
 CC A composite (incomplete) nucleotide sequence (AAT42753) and putative  
 CC translation (AAW08145) sequence are provided of a motif that is found  
 CC spliced to the 3' end of chicken CHD-W clone CC14 (see also AAT42754).  
 CC There are no splice donor or acceptor sites within the motif  
 CC suggesting it is a final rather than an intermediary product of  
 CC splicing. The motif is also found as an insert in some CHD-1A clones  
 CC (see also AAT42752).  
 CC  
 XX Sequence 137 BP; 75 A; 10 C; 34 G; 18 T; 0 other;  
 SQ  
 Query Match 12.5%; Score 33; DB 18; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 103 AGAAAG 135  
 Db 103 agaaag 135  
 RESULT 3  
 ABA21438/c  
 ID ABA21438 standard; DNA; 393 BP.  
 XX  
 AC ABA21438;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Human nervous system related polynucleotide SEQ ID NO 13769.  
 XX  
 KW Human; noctropic; neuroprotective; cytostatic; dermatological; virocidic;  
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antislaking; antianemic; antiarteritic; cancer;  
 KW antilemmatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiparasitic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 XX  
 PD 16-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01334.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.

PR	30-JUN-2000	2000US-0215135	PR	20-OCT-2000	2000US-0241808	PR	20-OCT-2000	2000US-0241809	PR	20-OCT-2000	2000US-0241809
PR	07-JUL-2000	2000US-0216647	PR	20-OCT-2000	2000US-0241826	PR	20-OCT-2000	2000US-0241826	PR	20-OCT-2000	2000US-0241826
PR	11-JUL-2000	2000US-0216880	PR	20-OCT-2000	2000US-0242221	PR	20-OCT-2000	2000US-0242221	PR	20-OCT-2000	2000US-0242221
PR	11-JUL-2000	2000US-0217487	PR	01-NOV-2000	2000US-0244617	PR	01-NOV-2000	2000US-0244617	PR	01-NOV-2000	2000US-0244617
PR	14-JUL-2000	2000US-0218290	PR	08-NOV-2000	2000US-0246474	PR	08-NOV-2000	2000US-0246474	PR	08-NOV-2000	2000US-0246474
PR	26-JUL-2000	2000US-0220963	PR	08-NOV-2000	2000US-0246475	PR	08-NOV-2000	2000US-0246475	PR	08-NOV-2000	2000US-0246475
PR	26-JUL-2000	2000US-0220964	PR	08-NOV-2000	2000US-0246476	PR	08-NOV-2000	2000US-0246476	PR	08-NOV-2000	2000US-0246476
PR	14-AUG-2000	2000US-0224518	PR	08-NOV-2000	2000US-0246477	PR	08-NOV-2000	2000US-0246477	PR	08-NOV-2000	2000US-0246477
PR	14-AUG-2000	2000US-0224519	PR	08-NOV-2000	2000US-0246478	PR	08-NOV-2000	2000US-0246478	PR	08-NOV-2000	2000US-0246478
PR	14-AUG-2000	2000US-0225213	PR	08-NOV-2000	2000US-0246523	PR	08-NOV-2000	2000US-0246523	PR	08-NOV-2000	2000US-0246523
PR	14-AUG-2000	2000US-0225214	PR	08-NOV-2000	2000US-0246524	PR	08-NOV-2000	2000US-0246524	PR	08-NOV-2000	2000US-0246524
PR	14-AUG-2000	2000US-0225266	PR	08-NOV-2000	2000US-0246525	PR	08-NOV-2000	2000US-0246525	PR	08-NOV-2000	2000US-0246525
PR	14-AUG-2000	2000US-0225267	PR	08-NOV-2000	2000US-0246526	PR	08-NOV-2000	2000US-0246526	PR	08-NOV-2000	2000US-0246526
PR	14-AUG-2000	2000US-0225268	PR	08-NOV-2000	2000US-0246527	PR	08-NOV-2000	2000US-0246527	PR	08-NOV-2000	2000US-0246527
PR	14-AUG-2000	2000US-0225270	PR	08-NOV-2000	2000US-0246528	PR	08-NOV-2000	2000US-0246528	PR	08-NOV-2000	2000US-0246528
PR	14-AUG-2000	2000US-0225271	PR	08-NOV-2000	2000US-0246532	PR	08-NOV-2000	2000US-0246532	PR	08-NOV-2000	2000US-0246532
PR	14-AUG-2000	2000US-0225757	PR	08-NOV-2000	2000US-0246610	PR	08-NOV-2000	2000US-0246610	PR	08-NOV-2000	2000US-0246610
PR	14-AUG-2000	2000US-0225758	PR	08-NOV-2000	2000US-0246611	PR	08-NOV-2000	2000US-0246611	PR	08-NOV-2000	2000US-0246611
PR	14-AUG-2000	2000US-0225759	PR	08-NOV-2000	2000US-0246613	PR	08-NOV-2000	2000US-0246613	PR	08-NOV-2000	2000US-0246613
PR	18-AUG-2000	2000US-0226279	PR	17-NOV-2000	2000US-0249207	PR	17-NOV-2000	2000US-0249207	PR	17-NOV-2000	2000US-0249207
PR	22-AUG-2000	2000US-0226861	PR	17-NOV-2000	2000US-0249208	PR	17-NOV-2000	2000US-0249208	PR	17-NOV-2000	2000US-0249208
PR	22-AUG-2000	2000US-0226868	PR	17-NOV-2000	2000US-0249209	PR	17-NOV-2000	2000US-0249209	PR	17-NOV-2000	2000US-0249209
PR	22-AUG-2000	2000US-0227182	PR	17-NOV-2000	2000US-0249210	PR	17-NOV-2000	2000US-0249210	PR	17-NOV-2000	2000US-0249210
PR	23-AUG-2000	2000US-0227009	PR	17-NOV-2000	2000US-0249211	PR	17-NOV-2000	2000US-0249211	PR	17-NOV-2000	2000US-0249211
PR	30-AUG-2000	2000US-0228924	PR	17-NOV-2000	2000US-0249212	PR	17-NOV-2000	2000US-0249212	PR	17-NOV-2000	2000US-0249212
PR	01-SEP-2000	2000US-0229287	PR	17-NOV-2000	2000US-0249213	PR	17-NOV-2000	2000US-0249213	PR	17	



```

RESULT 6
AAZ29254 ID AAZ29254 standard; DNA; 978 BP.
XX
XX
AC AAZ29254;
XX
XX 28-FEB-2000 (first entry)
XX
XX Rffin 3198 gene.
XX
XX Rffin 3198 gene: P.falciparum chromosome 2; Rffin protein;
XX mammalian expression plasmid VR1050; DNA vaccine; immunisation;
XX stage specific protein expression; antimicrobial vaccine;
XX antimicrobial drug; IAF; immunofluorescent antibody testing; ds.
XX
XX Plasmodium falciparum.
XX
XX MO955381-A1.
XX
XX 04-NOV-1999.
XX
XX 26-APR-1999; 99MO-US09047.
XX
XX 24-APR-1998; 98US-0082947.
XX
XX 23-APR-1999; 99US-0082947.
XX
XX (USNA ) US SEC OF NAVY.
XX
XX Hoffman SL, Carucci DJ;
XX
XX WPI; 2000-086380/07.
XX
XX Use of microbial, animal and/or human genomic data for identification
XX of vaccine, drug or diagnostic products -
XX
XX Claim 16; Page 17; 23pp; English.
XX
XX The present sequence is the Rffin 3198 gene of chromosome 2. Rffin genes
XX were used to characterise the protein expression from various life cycle
XX stages of P.falciparum. Oligonucleotide primers were used to amplify each
XX selected ORF from chromosome 2 using genomic DNA as template. The
XX purified amplified products were cloned into the mammalian expression
XX plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera
XX obtained from groups of mice immunised with doses of vaccine was used to
XX identify stage specific expression of Rffin protein by immunofluorescent
XX antibody testing. The method is useful for determination of subcellular
XX localisation of proteins and for the development of antimicrobial
XX vaccines and drugs.
XX
XX Sequence 978 BP; 357 A; 152 C; 208 G; 261 T; 0 other;
XX
XX
XX Query Match 8.3%; Score 22; DB 21; Length 978;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX OY 45 AGAAAAAGAAATGAAGAAAA 66
XX |||||||||||||||||||
XX Db 925 agaaaaaagaatgaagaanaa 946
XX
XX
XX RESULT 7
XX AAA70286 ID AAA70286 standard; DNA; 984 BP.
XX
XX AAA70286;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:419.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX

```

```

XX
XX OS Plasmodium falciparum.
XX
XX MO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99MO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF) HOFFMAN S.
XX (CARU) CARUCCI D.
XX (GARD) GARDNER M.
XX (VENT) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 575; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasite life cycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB1814 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 984 BP; 360 A; 145 C; 183 G; 296 T; 0 other;
XX
XX
XX Query Match 8.3%; Score 22; DB 21; Length 984;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX OY 45 AGAAAAAGAAATGAAGAAAA 66
XX |||||||||||||||||||
XX Db 931 agaaaaaagaatgaagaanaa 952
XX
XX
XX RESULT 8
XX AAA70284 ID AAA70284 standard; DNA; 1038 BP.
XX
XX AAA70284;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:417.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX

```

```
OS Plasmodium falciparum.
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 574-575; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially in the detection of infection with P. falciparum, membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 1038 BP; 379 A; 157 C; 209 G; 293 T; 0 other;
SQ
Query Match 8.3%; Score 22; DB 21; Length 1038;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 AGAAAAAGAAATGAGGAAAA 66
Db 985 agaaaaaagaatatgaagaaaa 1006
RESULT 9
AAA70279
ID AAA70279 standard; DNA; 1047 BP.
XX
XX AAA70279;
AC
XX
XX 07-NOV-2000 (first entry)
DT
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:412.
DE
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
KM
XX Plasmodium falciparum.
OS
```

```
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 573; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially in the detection of infection with P. falciparum, membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 1047 BP; 383 A; 163 C; 211 G; 290 T; 0 other;
SQ
Query Match 8.3%; Score 22; DB 21; Length 1047;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 AGAAAAAGAAATGAGGAAAA 66
Db 994 agaaaaaagaatatgaagaaaa 1015
RESULT 10
ABL18816
ID ABL18816 standard; DNA; 29783 BP.
XX
XX ABL18816;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 7921.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
KM
XX Drosophila melanogaster.
OS
```

PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001MO-US09231.  
PE  
XX 23-MAR-2000; 2000US-191637P.  
FR  
XX 11-JUL-2000; 2000US-0614150.  
PA  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EM;  
PI WPI: 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PR interactions -  
XX  
XX Claim 1; SEQ ID NO 7921; 21np + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 29783 BP; 7503 A; 6790 C; 6798 G; 8692 T; 0 other;

Query Match 8.3%; Score 22; DB 23; Length 29783;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 AAAAAATATAAAACGAAAAA 52  
DB 19930 aaaaaataaaacagaaaaa 19951

RESULT 11  
AA181813/C  
ID AA181813 standard; cDNA; 471 BP.  
XX  
XX AA181813;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polynucleotide SEQ ID NO 1873.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001MO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX 18-MAY-2000; 2000US-0577409.  
PA  
XX (HYSE-) HYSEQ INC.  
XX

PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-514838/56.  
DR  
XX P-PSDB; AA001882.  
DR  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PR disorders -  
XX  
XX Claim 1; SEQ ID NO 1873; 1399pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 471 BP; 121 A; 123 C; 126 G; 97 T; 4 other;

Query Match 7.9%; Score 21; DB 22; Length 471;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 146 AAAAAAGGAGAAAAAAGAAAA 166  
DB 73 AAAAAAGGAGAAAAAAGAAAA 53

RESULT 12  
AA144389  
ID AA144389 standard; DNA; 588 BP.  
XX  
XX AA144389;  
AC  
XX  
XX 17-OCT-2001 (first entry)  
DT  
XX  
XX Probe #13075 used to measure gene expression in human placenta sample.  
DE  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157272-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001MO-US00663.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
XX 26-MAY-2000; 2000US-0207456.  
PR  
XX 30-JUN-2000; 2000US-0608408.  
PR  
XX 03-AUG-2000; 2000US-0632366.  
PR  
XX 21-SEP-2000; 2000US-0234687.  
PR  
XX 27-SEP-2000; 2000US-0236359.  
PR  
XX 04-OCT-2000; 2000GB-0024263.  
PA  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI: 2001-488897/53.  
XX  
XX

PT Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human placenta -  
 XX Claim 25; SEQ ID No 13075; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 XX  
 SQ Sequence 588 BP; 280 A; 84 C; 115 G; 109 T; 0 other;

Query Match 7.9%; Score 21; DB 22; Length 588;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 172 TTAGCAAGCAATTAAAGAAA 192  
 ||||||||||||||||||  
 Db 218 taagaagaataaagaataa 238

## RESULT 13

AA229256  
 ID AA229256 standard; DNA; 936 BP.

XX  
 AC AA229256;

XX  
 DT 28-FEB-2000 (first entry)

XX  
 DE Rifin 3201 gene.

XX  
 KM Rifin 3201 gene; P.falciparum chromosome 2; Rifin protein;

KM mammalian expression plasmid VR1050; DNA vaccine; immunisation;

KM stage specific protein expression; antimicrobial vaccine;

KM antimicrobial drug; IAF; immunofluorescent antibody testing; ds.

XX  
 OS Plasmodium falciparum.

XX  
 PN WO955381-A1.

XX  
 PD 04-NOV-1999.

XX  
 PF 26-APR-1999; 99WO-US09047.

XX  
 PR 24-APR-1998; 98US-0082947.

XX  
 PR 23-APR-1999; 99US-0082947.

XX  
 PA (USNA ) US SEC OF NAVY.

XX  
 PI Hoffman SL, Carucci DJ;

XX  
 DR WPI; 2000-086380/07.

XX  
 PT Use of microbial, animal and/or human genomic data for identification  
 PT of vaccine, drug or diagnostic products -

XX  
 PS Claim 18; Page 17; 23pp; English.

XX  
 CC The present sequence is the Rifin 3201 gene of chromosome 2. Rifin genes  
 CC were used to characterise the protein expression from various life cycle  
 CC stages of P.falciparum. Oligonucleotide primers were used to amplify each  
 CC selected ORF from chromosome 2 using genomic DNA as template. The

XX  
 CC purified amplified products were cloned into the mammalian expression  
 CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera

XX  
 CC obtained from groups of mice immunised with doses of vaccine was used to  
 CC identify stage specific expression of Rifin protein by immunofluorescent  
 CC antibody testing. The method is useful for determination of subcellular

XX  
 CC localisation of proteins and for the development of antimicrobial  
 CC vaccines and drugs.

XX  
 SQ Sequence 936 BP; 357 A; 136 C; 190 G; 253 T; 0 other;

Query Match 7.9%; Score 21; DB 21; Length 936;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 GAAAAAGAAATGAGAAAA 66  
 ||||||||||||||||||  
 Db 884 gaaaagaataagaataa 904

## RESULT 14

AAH33151  
 ID AAH33151 standard; CDNA; 976 BP.

XX  
 AC AAH33151;

XX  
 DT 03-SEP-2001 (first entry)

XX  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:207.

XX  
 KM Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX  
 KM colorectal carcinoma; chromosome 16; ss.

XX  
 OS Homo sapiens.

XX  
 PN WO200122920-A2.

XX  
 PD 05-APR-2001.

XX  
 PF 28-SEP-2000; 2000WO-US26524.

XX  
 PR 29-SEP-1999; 99US-0157137.

XX  
 PR 03-NOV-1999; 99US-0163280.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX  
 DR WPI; 2001-235357/24.

XX  
 DR P-PSDB; AAG73720.

XX  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX  
 PS Claim 1; Page 2352; 9803pp; English.

XX  
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where

XX  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene

XX  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P

XX  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions

XX  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.

XX  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell

XX  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

XX  
 CC present invention.

XX  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for

XX  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX  
 SQ Sequence 976 BP; 431 A; 164 C; 204 G; 168 T; 9 other;

Query Match 7.9%; Score 21; DB 22; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:57 : Search time 7016.61 Seconds

(without alignments)  
509,747 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265

Sequence: 1 GATGAGATGTCTTCACTGAA.....AAGAGTGAAGAGAGAGAG 265

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthm: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: qb\_estl: \*  
10: qb\_est2: \*  
11: qb\_hic: \*  
12: qb\_gss: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_hiv: \*  
15: em\_gss\_pln: \*  
16: em\_gss\_vrt: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	10.2	882	12	AZ184391 SP.1002.B
2	26	9.8	590	12	AZ091555 RPT-23-2
3	24	9.1	343	9	AA748563 oa57c09.s
4	24	9.1	375	12	AZ213624 Sheared D
5	24	9.1	430	9	A1242163 qh81f02.x
6	24	9.1	456	9	AA699918 z161f12.s
7	24	9.1	485	12	AO642393 RPT193-DP
8	24	9.1	539	12	AO500417 v40b12.mf
9	24	9.1	643	12	AO952151 Sheared D
10	24	9.1	657	12	AZ508184 LM0350318
11	24	9.1	677	10	BM015587 BM015587
12	24	9.1	821	10	BF239967 601905170
13	23	8.7	372	9	AU039357 AU039357
14	23	8.7	393	9	AA658426 nui8e11.s
15	23	8.7	683	9	AA653205 AL653205
16	23	8.7	695	10	B1640191 SD23001.5
17	23	8.7	743	12	AG032665 Pan trogl

C 18	23	8.7	826	12	AG912123	AG912123 nbe00171
C 19	23	8.7	867	12	AZ668023	AZ668023 ENTJ807F
C 20	23	8.7	920	12	AZ674814	AZ674814 ENTJW787R
C 21	22	8.3	154	12	AG059567	AG059567 Pan trogl
C 22	22	8.3	301	9	A1479119	A1479119 tm31e03.x
C 23	22	8.3	413	12	AZ161843	AZ161843 SP.0070.B
C 24	22	8.3	415	10	BM274338	BM274338 PEST00a4
C 25	22	8.3	518	12	TA169E06P	TA169E06 T. brucei
C 26	22	8.3	520	10	BM273317	BM273317 PEST00a4
C 27	22	8.3	554	10	BM275208	BM275208 PEST00a6
C 28	22	8.3	560	10	BM274989	BM274989 PEST00a7
C 29	22	8.3	584	10	BM273877	BM273877 PEST00a6
C 30	22	8.3	616	12	AA550018	AA550018 1097m3.gm
C 31	22	8.3	619	12	AZ831967	AZ831967 2M0112G05
C 32	22	8.3	876	12	AZ539395	AZ539395 ENPPE09TF
C 33	22	8.3	884	9	AU067771	AU067771 AU067771
C 34	22	8.3	938	12	CNS01SN3	AL165432 Tetradon
C 35	22	8.3	950	10	BE961679	BE961679 601647954
C 36	22	8.3	980	12	CNS0450L	AL275646 Tetradon
C 37	22	8.3	1062	12	CNS02MNV	AL205168 Tetradon
C 38	22	8.3	1248	10	BC851122	BC851122 1024030D0
C 39	21	7.9	148	10	BR075197	BR075197 224037.WA
C 40	21	7.9	165	9	AM368791	AM368791 IL27H019
C 41	21	7.9	208	9	A1652370	A1652370 WB06908.x
C 42	21	7.9	254	9	AA355716	AA355716 EST64158
C 43	21	7.9	273	12	AG059338	AG059338 Pan trogl
C 44	21	7.9	310	9	A1033270	A1033270 ow/4a06.s
C 45	21	7.9	355	9	AA305433	AA305433 EST176411

#### ALIGNMENTS

RESULT 1  
LOCUS AZ184391  
DEFINITION SP.1002.B2.G05.T7A Strongylocentrotus purpuratus, purple sea urchin  
clone Plate-1002 Col-10 Row-N, DNA sequence.

ACCESSION AZ184391  
VERSION AZ184391.1 GI:8356766

SOURCE  
KEYWORDS GSS.  
ORGANISM Strongylocentrotus purpuratus.

REFERENCE  
AUTHORS Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and Hood,L.

TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)  
MEDLINE 20402566

COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 1002 row: N column: 10  
Seq primer: 77  
Class: BAC ends  
High quality sequence stop: 882.

FEATURES  
source  
1..882  
Location/Qualifiers  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="Plate-1002 Col-10 Row-N"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea





Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
Seq primer: -40m13 fwd. ER from Amersham  
High quality sequence stop: 420.

## FEATURES

source

```
1. .456
/organism="Homo sapiens"
/db_xref="GDB:1335080"
/db_xref="taxon:9606"
/clone="IMAGE:435311"
/clone_id="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer (5'
ACGTGAGAGATTAAATTAAGATCTTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

## BASE COUNT

119 a 95 c 57 g 185 t

## ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 50;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AAAAGAAAAAGAGTAAAGAGA 180

Db 47 AAAAGAAAAAGAGATAGAGAGA 24

## RESULT 7

LOCUS AO642393 485 bp DNA linear GSS 08-JUL-1999  
DEFINITION RCI193-DpniI-28K15.TV RCI193-DpniI Trypanosoma brucei genomic clone  
ACCESSION AO642393  
VERSION AO642393.1 GI:5119103  
KEYWORDS GSS.

SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 485)  
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,  
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,  
Fraser,C. and Adams,M.

TITLE Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RCI-93  
JOURNAL library for gene discovery and sequence-ready map construction  
COMMENT Other\_GSSs: RCI193-DpniI-28K15.TV  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

FEATURES  
Email: neisayed@tigr.org  
Clones and high density filters may be purchased from BACPAC  
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search  
page: http://www.tigr.org/tdb/mdb/tbdb/.  
Seq primer: T7  
Class: BAC ends.

Location/Qualifiers

## source

```
1. .485
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-DpniI-28K15"
/clone_id="RPC193-DpniI"
/notes="Vector: pBACE3.6; Site_1: Bam HI; Site_2: Bam HI;
constructed for The Institute for Genomic Research by
Boni Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methylase (RPC193-EcoRI segment) or Dpn II (RPC193-DpniI
segment). High molecular weight fragments were ligated in
pBACE3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 x the haploid
non-mitochondrial genome."
```

## BASE COUNT

172 a 91 c 68 g 153 t 1 others

## ORIGIN

Query Match 9.1%; Score 24; DB 12; Length 485;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 TAAGAAAGATTAAAGAGAGA 195

Db 64 TAAGAAAGATTAAAGAGAGA 87

## RESULT 8

AO500417/c 539 bp DNA linear GSS 29-APR-1999  
LOCUS AO500417  
DEFINITION V40B12 mtH-3xHA/lacZ Insertion Library Saccharomyces cerevisiae  
ACCESSION AO500417  
VERSION AO500417.1 GI:4706127  
KEYWORDS GSS.

## SOURCE

ORGANISM baker's yeast.  
Saccharomyces cerevisiae  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 539)  
AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,  
desEtages,S.A., Chung,K.-H., Sheehan,A., Symonistis,D., Jansen,R.,  
Umansky,L., Heidman,K., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,  
Hager,K., Miller,P., Roeder,G.S. and Snyder,M.

TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and  
JOURNAL Gene Disruption  
COMMENT Unpublished (1999)  
Contact: Kumar A

## JOURNAL

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology  
Yale University  
P.O. Box 208103, New Haven, CT 06520-8103, USA  
Tel: 203 432 9949  
Fax: 203 432 6161

FEATURES  
Email: anuj.kumar@yale.edu  
te of mtH-3xHA/lacZ insertion.  
Seq primer: GGCTCTCTCTTGGAGATGAC  
Class: transposon-tagged.  
Location/Qualifiers

## source

```
1. .539
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_id="mtH-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/notes="Vector: pHS56-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mtH-3xHA/lacZ
mitransposon containing lacZ, URA3, and tet resistance."
```

BASE COUNT 58 a 142 c 24 g 315 t  
ORIGIN

Query Match 9.1%; Score 24; DB 12; Length 539;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 145 GAAAGGAGAGAAAAAGAAAAGA 168  
|||||

Db 257 GAAAGGAGAGAAAAAGAAAAGA 234

RESULT 9  
A0952151 643 bp DNA linear GSS 27-JAN-2000  
LOCUS Sheared DNA-42C8.TF Sheared DNA Trypanosoma brucei genomic clone  
DEFINITION Sheared DNA-42C8, DNA sequence.  
ACCESSION A0952151  
VERSION A0952151.1 GI:6775416  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 643)  
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
Fraser, C. and Adams, M.  
Determination of clone end sequences from Trypanosoma brucei GUTat  
10.1 sheared DNA library  
Unpublished (1999)  
Other\_GSSs: Sheared DNA-42C8.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
DNA library constructed at TIGR. Clones will be available for  
distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tdb/mdb/tbdb/>.  
Seq primer: M13-Forward  
Class: shotgun.  
FEATURES  
Location/Qualifiers  
source 1..643  
/organism="Trypanosoma brucei"  
/strain="TRE927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-42C8"  
/clone\_lib="Sheared DNA"  
/note="Vector: pUC18; Site: 1: SmaI; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TRE927/4 GUTat 10.1) was mechanically  
sheared to give a tight size distribution (approx 2 kb).  
The v + 1 method used for the library construction is  
described in detail in Smith, H.O. and Venter, J.C.  
(Making small insert libraries for whole genome shotgun  
sequencing projects. In Genome Sequencing: A Practical  
Approach, eds. M. Vaudin and B. Barrell, Oxford University  
Press, 1999)."

BASE COUNT 188 a 135 c 111 g 208 t 1 others  
ORIGIN

Query Match 9.1%; Score 24; DB 12; Length 643;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 TAAGAAAGATTAAGAAAAAGA 195

Db 172 TAAGAAAGATTAAGAAAAAGA 149  
|||||

RESULT 10  
A2508184/c 657 bp DNA linear GSS 05-OCT-2000  
LOCUS clone U0GC1M0350G18 F, DNA sequence.  
DEFINITION clone U0GC1M0350G18 F, DNA sequence.  
ACCESSION A2508184  
VERSION A2508184.1 GI:10689500  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 657)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0350 row: G column: 18  
Seq primer: CGTTGTAACACGACGCCACT  
Class: plasmid ends  
High quality sequence stop: 657.  
FEATURES  
Location/Qualifiers  
source 1..657  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U0GC1M0350G18"  
/clone\_lib="Mouse 10kb plasmid library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD42hy; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (g11473214[90]AF129072.1), a copy number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 182 a 156 c 84 g 235 t  
ORIGIN

Query Match 9.1%; Score 24; DB 12; Length 657;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 ATAAAAAAATATAACAGAAAAA 52  
 Db 298 ATAAAAAAATATAACAGAAAAA 275

RESULT 11  
 BM015597 677 bp mRNA linear EST 30-OCT-2001  
 LOCUS 603641828F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5417887 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BM015597 GI:16529951  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 677)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM12065 row: 9 column: 08  
 High quality sequence stop: 674.  
 Location/Qualifiers

FEATURES  
 source  
 1. 677  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5417887"  
 /clone\_lib="NIH\_MGC\_87"  
 /tissue\_type="mammary adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: breast; Vector: pCMV-SPOK6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; Oligo-dT primed.  
 Average insert size 1.383 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

BASE COUNT 305 a 92 c 153 g 127 t  
 ORIGIN

Query Match 9.1%; Score 24; DB 10; Length 677;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAAAGAGATAGAAAGA 180  
 Db 245 AAAAGAAAAAGAGATAGAAAGA 268

RESULT 12  
 BF239967 821 bp mRNA linear EST 14-NOV-2000  
 LOCUS 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 821)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers

FEATURES  
 source  
 1. 821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4133129"  
 /clone\_lib="NIH\_MGC\_54"  
 /tissue\_type="from chronic myelogenous leukemia"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccataggcc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGGCCATATGCGC-3' and 3' adaptor  
 sequence: 5'-ATCTAGAGCGCGCGCGCGCAG-3' (30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN

Query Match 9.1%; Score 24; DB 10; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAAAGAGATAGAAAGA 180  
 Db 391 AAAAGAAAAAGAGATAGAAAGA 414

RESULT 13  
 AU039357 372 bp mRNA linear EST 29-MAR-1999  
 LOCUS AU039357 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
 DEFINITION AU039357 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
 ACCESSION AU039357  
 VERSION AU039357.1 GI:4008597  
 KEYWORDS EST.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 REFERENCE 1 (bases 1 to 372)  
 Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
 Yoshino,R., Mittra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,  
 Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
 The Dictyostelium developmental cDNA project: generation and  
 analysis of expressed sequence tags from the first-finger stage of  
 development  
 DNA Res. 5 (6), 335-340 (1998)  
 MEDLINE 99156227  
 COMMENT Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
 Email: d402h@tsukuba.ac.jp  
 PROJECT = "Dictyostelium discoideum cDNA project in Japan".  
 Location/Qualifiers

FEATURES  
 source  
 1. 372  
 /organism="Dictyostelium discoideum"



/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLH629"  
/clone\_1lb="Dictyostellium discoideum SL (H.Drushihara)"  
/dev\_stage="sing"  
BASE COUNT 234 a 40 c 51 g 46 t 1 others  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 AAGAGAAAGATATTAAGAAAA 207  
|||||  
DB 252 AAGAGAAAGATATTAAGAAAA 274

RESULT 14  
AA658426 393 bp mRNA linear EST 05-NOV-1997  
LOCUS AA658426  
DEFINITION nubl1.51 NCI\_CGAP\_Pr2 Homo sapiens cDNA clone IMAGE:1208396, mRNA  
sequence.  
ACCESSION AA658426  
VERSION AA658426.1 GI:2594580  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 393)  
NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaps-rt@mail.nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuqui, M.D.  
Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bdnp/image/image.html  
Seq primer: -40m13 fwd. ET from Amerisham.  
Location/Qualifiers

FEATURES  
source  
1. 393  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1208396"  
/clone\_1lb="NCI\_CGAP\_Pr2"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site.1: NotI; Site.2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated total cellular RNA obtained from 5,000-10,  
000 microdissected preneoplastic cells  
histologically-determined to be prostatic intraepithelial  
neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Kitzman."  
BASE COUNT 121 a 58 c 69 g 145 t  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 41 AAGAGAAAGAAAGAAAGTGAAGA 63  
|||||  
DB 391 AAGAGAAAGAAAGAAAGTGAAGA 369

RESULT 15  
AL653205 683 bp mRNA linear EST 13-DEC-2001  
LOCUS AL653205  
DEFINITION AL653205 XGC-gastrula silurana tropicalis cDNA clone Tgas049n16 5',  
mRNA sequence.  
ACCESSION AL653205  
VERSION AL653205.1 GI:17664069  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;  
Xenopodinae; Silurana.  
Xenopus laevis; Silurana.  
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
1 (bases 1 to 683)  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
Unpublished (2001)  
Contact: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: Tgas049n16.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
Location/Qualifiers

FEATURES  
source  
1. 683  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="Tgas049n16"  
/clone\_1lb="XGC-gastrula"  
/dev\_stage="gastrula (stages 10.5-13 mixed)"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA  
was oligo dT primed from sug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."  
BASE COUNT 225 a 113 c 110 g 234 t 1 others  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 AATATTAACAGAAAAAGAAAA 57  
|||||  
DB 517 AATATTAACAGAAAAAGAAAA 539

Search completed: August 2, 2002, 22:42:01  
Job time: 30176 sec

---





TITLE  
JOURNAL  
COMMENT

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,  
Sever, P., Spencer, B., Stange, Thomas, N., Stojanovic, N.,  
Strauss, N., Sudriaman, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trifilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIMR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L20703

Center clone name: 438\_D\_3

\* NOTE: This record contains 77 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 729 828: contig of 728 bp in length  
\* 829 1553: contig of 725 bp in length  
\* 1554 1653: gap of 100 bp  
\* 1654 2373: contig of 720 bp in length  
\* 2374 2473: gap of 100 bp  
\* 2474 3176: contig of 703 bp in length  
\* 3177 3276: gap of 100 bp  
\* 3277 3998: contig of 722 bp in length  
\* 3999 4098: gap of 100 bp  
\* 4099 4821: contig of 723 bp in length  
\* 4822 4921: gap of 100 bp  
\* 4922 5607: contig of 686 bp in length  
\* 5608 5707: gap of 100 bp  
\* 5708 6414: contig of 707 bp in length  
\* 6415 6514: gap of 100 bp  
\* 6515 7238: contig of 724 bp in length  
\* 7239 7338: gap of 100 bp  
\* 7339 8067: contig of 729 bp in length  
\* 8068 8167: gap of 100 bp  
\* 8168 8898: contig of 731 bp in length  
\* 8899 8998: gap of 100 bp  
\* 8999 9728: contig of 730 bp in length  
\* 9729 9828: gap of 100 bp  
\* 9829 10554: contig of 726 bp in length  
\* 10555 10654: gap of 100 bp  
\* 10655 11300: contig of 646 bp in length  
\* 11301 11400: gap of 100 bp  
\* 11401 12105: contig of 705 bp in length  
\* 12106 12205: gap of 100 bp  
\* 12206 12867: contig of 662 bp in length  
\* 12868 12967: gap of 100 bp  
\* 12968 13666: contig of 699 bp in length  
\* 13667 13766: gap of 100 bp  
\* 13767 14494: contig of 728 bp in length  
\* 14495 14594: gap of 100 bp  
\* 14595 15328: contig of 734 bp in length  
\* 15329 15428: gap of 100 bp

15429 16122: contig of 694 bp in length  
\* 16123 16222: gap of 100 bp  
\* 16223 16944: contig of 722 bp in length  
\* 16945 17044: gap of 100 bp  
\* 17045 17738: contig of 694 bp in length  
\* 17739 17838: gap of 100 bp  
\* 17839 18534: contig of 696 bp in length  
\* 18535 18634: gap of 100 bp  
\* 18635 19333: contig of 699 bp in length  
\* 19334 19433: gap of 100 bp  
\* 19434 20145: contig of 712 bp in length  
\* 20146 20245: gap of 100 bp  
\* 20246 20972: contig of 727 bp in length  
\* 20973 21072: gap of 100 bp  
\* 21073 21805: contig of 733 bp in length  
\* 21806 21905: gap of 100 bp  
\* 21906 22639: contig of 734 bp in length  
\* 22640 22739: gap of 100 bp  
\* 22740 23460: contig of 721 bp in length  
\* 23461 23560: gap of 100 bp  
\* 23561 24284: contig of 724 bp in length  
\* 24285 24384: gap of 100 bp  
\* 24385 25099: contig of 715 bp in length  
\* 25100 25199: gap of 100 bp  
\* 25200 25902: contig of 703 bp in length  
\* 25903 26002: gap of 100 bp  
\* 26003 26698: contig of 696 bp in length  
\* 26699 26798: gap of 100 bp  
\* 26799 27503: contig of 705 bp in length  
\* 27504 27603: gap of 100 bp  
\* 27604 28337: contig of 734 bp in length  
\* 28338 28437: gap of 100 bp  
\* 28438 29161: contig of 724 bp in length  
\* 29162 29261: gap of 100 bp  
\* 29262 29984: contig of 723 bp in length  
\* 29985 30084: gap of 100 bp  
\* 30085 30784: contig of 700 bp in length  
\* 30785 30884: gap of 100 bp  
\* 30885 31509: contig of 625 bp in length  
\* 31510 31609: gap of 100 bp  
\* 31610 32316: contig of 707 bp in length  
\* 32317 32416: gap of 100 bp  
\* 32417 33135: contig of 719 bp in length  
\* 33136 33235: gap of 100 bp  
\* 33236 33944: contig of 709 bp in length  
\* 33945 34044: gap of 100 bp  
\* 34045 34766: contig of 722 bp in length  
\* 34767 34866: gap of 100 bp  
\* 34867 35351: contig of 685 bp in length  
\* 35352 35651: gap of 100 bp  
\* 35652 36388: contig of 737 bp in length  
\* 36389 36488: gap of 100 bp  
\* 36489 37208: contig of 720 bp in length  
\* 37209 37308: gap of 100 bp  
\* 37309 38035: contig of 727 bp in length  
\* 38036 38135: gap of 100 bp  
\* 38136 38858: contig of 723 bp in length  
\* 38859 38958: gap of 100 bp  
\* 38959 39684: contig of 726 bp in length  
\* 39685 39784: gap of 100 bp  
\* 39785 40482: contig of 698 bp in length  
\* 40483 40582: gap of 100 bp  
\* 40583 41286: contig of 704 bp in length  
\* 41287 41386: gap of 100 bp  
\* 41387 42090: contig of 704 bp in length  
\* 42091 42190: gap of 100 bp  
\* 42191 42911: contig of 721 bp in length  
\* 42912 43011: gap of 100 bp  
\* 43012 43740: contig of 729 bp in length  
\* 43741 43840: gap of 100 bp  
\* 43841 44567: contig of 727 bp in length  
\* 44568 44667: gap of 100 bp  
\* 44668 45371: contig of 704 bp in length



/translation="NPISATVYCALLIGLGTOMYSNLRKIPYMRHPEHYVGMGLGAV  
FANOLVKMDVLTKEEDLDVNLAKARANDERRRYFQIYTPILVYGLSASQILLFPRKIGDV  
FCV"  
complement(19438..19557)  
/gene="F1C12.70"  
/number=1  
complement(19558..19648)  
/gene="F1C12.70"  
/number=2  
21131..22045  
/gene="F1C12.80"  
/number=1  
21131..25119  
/gene="F1C12.80"  
join(21131..22045,22134..22261,22328..24208,24317..24665,  
24743..24890,24974..25119)  
/gene="F1C12.80"  
/note="similarity to mature-parasite-infected erythrocyte  
surface antigen MESA, Plasmodium falciparum, PIR2:A45605"  
/codon\_start=1  
/product="glu-rich protein"  
/protein\_id="CA18241.1"  
/db\_xref="GI:2982433"  
/db\_xref="SPTREMBL:O65430"  
/translation="MASSPEVMSQGFNCVLSRPHRCOTORDNTNNKKNNNNN  
VNTHVAFKKNLNVFPHLNCSSVVDVNDSTKAPPCSSKSSVSDRIKSKTE  
ESROVPSSSSLGSDSPNGOPNPSRNAASIVQIHPAPPTQOPSPNSILDSR  
TSSKSNLRNSESLESSEKESOMDIERNNEEEERIEGAPPELESQKREGRVY  
MOTIRKINSDETTTNNSSGNDNSKEVQTTTAKSPVOACSPRIKQHOLADLY  
QMTREKDLACLAKENYKSKFTNKGKIOSTLRIETRECLATGRRKRSKSDS  
NRSKSGVHMLHETKANSBETITGASTSHSTRIGRMORPHKATEKVKVLE  
KSGVLEKEVSAENAKGVLSLSEKNGKLKLEGTQRETVEGKREKTEK  
ISLESLEKGGKREKSTKAKLAKESVAKLVLIAKVLNNSSEKNNKRRKREK  
EGNVTWERKNDVLEATRRLLVESAESRSTTSKETKTRCAVAKVKKKED  
VLSLEKSKKEVDGNIKPGQVTTQDSCLRPAIEKVKQETVAKSOLKPEKRSR  
ERRLEIEEESTISGIAEKVAVNLSKKNRKAEMEGKGEKGAITNELLOASVEA  
KSKVDIEDKNNPQAVIHGSKERDKERNSSQNGEETLNRKSKSTKEIERQVTOE  
EKSVSHGKREKDKERNSSQNGEETLNRKSKSTKEIERQVTOE  
ESAKENSSQODESTHNPMDKKGIKEPEDESKKEREETENVEEVSVEFNDY  
DONEMEEDEEEEGDYFNGDDWDHDSIREPSYMEDEKREYLEVLTSEKDKCNL  
IERTVSNFLTSLDRKIDNLTITVOSHIGVLANOIEGDEYEWEVCSAPLEPD  
NTEEEEPKNTLEAPSDVCSOASSTMSMNFRRDODIDNDENPTTSLSEPLVPT  
NOSTODMOTISDLKEMOLOREMELERNTVKSCIMDIOKSVTODLSRSSSAAEQ  
RDPPKDKLKKCCVCEMPVDLSILYRFCLYIRNKSPTLK"

exon  
/gene="F1C12.80"  
/number=1  
22134..22261  
/gene="F1C12.80"  
/number=2  
22262..22327  
/gene="F1C12.80"  
/number=2  
22328..24208  
/gene="F1C12.80"  
/number=3  
24209..24316  
/gene="F1C12.80"  
/number=3  
24317..24665  
/gene="F1C12.80"  
/number=4  
24666..24742  
/gene="F1C12.80"  
/number=4  
24743..24890  
/gene="F1C12.80"  
/number=5  
24891..24973  
/gene="F1C12.80"  
/number=5  
24974..25119

intron  
exon

repeat\_region  
exon  
gene  
CDS  
/gene="F1C12.80"  
/number=6  
26862..26897  
/note="36bp TTC tandem repeat"  
26958..28126  
/gene="F1C12.90"  
/number=1  
26958..28942  
/gene="F1C12.90"  
join(26958..28126,28597..28942)  
/gene="F1C12.90"  
/note="similarity to gene F4P9.34 chromosome II BAC F4P9,  
Arabidopsis thaliana  
contains EST gb:237280, 237281"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA18242.1"  
/db\_xref="GI:2982434"  
/db\_xref="SPTREMBL:O65431"  
/translation="MAVVKREKONTFRKDLVGYINFSAEKLTLPALLVLTATL  
LPIPSFSLSTSDPFRCSIRFSASVPLNTTIVYESSSPSPKRNLDVLDGVTKR  
TFPTGSAAYNFMASVAGVNSFAVIGLSKRLHYVGHPSRCRWVSLDPTODIS  
TTFGTSLTDMGCGRLTYTVVNCPTSSISAVNPQNSGGLILHAATGDPILNLTDSIS  
VLNPEKSYDPLVNSRTKRYKDYLVGSSISXGNISPORVAVHVRERBSHP  
VLNAGCIHEVFEVILKRWTEIGRVTLIDIDPRDPCYVHNOPIYNDLHVRKMT  
KMFPEIDDEFLHVPKERTISSMESLEYSQFTIDKMSRRCYSGDPPATRYKM

alignment\_scores  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x ATF1C12 ..  
Align seg 1/1 to: ATF1C12 from: 1 to: 111945  
34 GYGLULYSARGLUThrLysGluLys 42  
|||||  
22605 GCGCAAAAGCGAGACGACGAGAAA 22631

seq\_name: gb\_hhg:AC094613  
seq\_documentation\_block:  
LOCUS AC094613 173449 bp DNA linear HTG 20-DEC-2001  
DEFINITION Rattus norvegicus clone CH230-5b2, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
66 unordered pieces.  
ACCESSION AC094613  
VERSION AC094613.2 GI:17941384  
KEYWORDS HTG; HNGS; PHASEL.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 173449)  
REFERENCE  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,  
Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,J.,  
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Butch,P., Butkelt,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhury,I., Christopoulos,C., Cleaveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Diaper,H.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Gara,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,R.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollaway,C.,  
Hollins,B., Honsi,F., Howard,S., Huber,J., Huljk,S., Hume,J.,







KIEFEMHSKVSMSWILGLCPISGLVNPVIGLIVYTELASGVPIHESFHTTNLPAIHVS  
VFEVVCVYLPVTVPEEERFLVNRIGPKTFMRFCYARGYKDLHKKDDFEFKLLTK  
LSFFIETEMPESTNSSTYSYNKHTQDSYVDLIHNNNNHNNMMDMFSSMDY  
TVSTLDIVASAEHLNHTVSPSDNTVEEETDELEFLKTCESGVVHIMGNTVVKART  
GSLPKKIAIDYVYALAKTRANSVILNHPHETLNLNGEVV"

intron /gene="AT4g19960"  
exon /number=1  
11621. 11846  
/gene="AT4g19960"  
intron /number=2  
11847. 11924  
/gene="AT4g19960"  
exon /number=3  
11925. 12170  
/gene="AT4g19960"  
intron /number=4  
12171. 12257  
/gene="AT4g19960"  
exon /number=5  
12258. 12332  
/gene="AT4g19960"  
intron /number=6  
12333. 12382  
/gene="AT4g19960"  
exon /number=7  
12383. 12643  
/gene="AT4g19960"  
intron /number=8  
12644. 12686  
/gene="AT4g19960"  
exon /number=9  
12687. 12887  
/gene="AT4g19960"  
intron /number=10  
12888. 12985  
/gene="AT4g19960"  
exon /number=11  
12986. 13288  
/gene="AT4g19960"  
intron /number=12  
13289. 13406  
/gene="AT4g19960"  
exon /number=13  
13407. 13929  
/gene="AT4g19960"  
intron /number=14  
13930. 14023  
/gene="AT4g19960"  
exon /number=15  
14024. 14567  
/gene="AT4g19960"  
intron /number=16  
15812. 22913  
/gene="AT4g19970"  
CDS join(15812. 16087,16379. 16756,17694. 18160,21289. 21472,  
21612. 21977,22437. 22913)  
/gene="AT4g19970"  
/note="similarity to predicted protein, Arabidopsis  
thaliana, PARCHX:E326963  
Contains Crystallins beta and gamma 'Greek key' motif  
signature AA308-323;Crystallins beta and gamma 'Greek key'  
motif signature AA647-662;Prokaryotic membrane lipoprotein  
lipid attachment site AA47-57;Prokaryotic membrane  
lipoprotein lipid attachment site AA390-400  
contains EST gb:AI993954.1, T42200"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB78997.1"  
/db\_xref="GI:7268792"  
/translation="MKSTNDESPSLIGYFAGVSIKDKSKPPAPVICYOGLRGRDVI  
RVLVLTIVTSLCLFYKSKNNPLNPFSPMKTDICYASKITNNSSSKTEPKPVSIL  
ERVLMNAAMEDNTVITLALNQAAEPNSTFDVRESFKVGIETERLLKVIAYCLDIK

ADQCCKVHPHCYLINATSDSOLSGFNRETPGYLKLIMRRDILROVYGLGYNFTFT  
DADIILRDPFPFPFPDADFOITCDYDNGSPDKKHHVNSGFYVYVANKTSKFKYK  
IRSRPEPKHDDPFIKNDIHEKLGIKMFEPTVYEGFCQSPRIINVNTMHA  
NCCIGLDNRYNNKALLEMKRYVSLNTVYSEFKMNI PPSFLDGSAGIOKEVK TLV  
LVYGLACILLYETAYLPDHOEDVNNLSRPLIDHSSSPSLRKSISFRYELNAS  
TENRTVIYVTTLNQAAEPNSLFLFLESFRIGGTCKLLOHVVYVCLDSKAPRCSOL  
HPNCYVLTGTTGDFSEKLEFAPDYLLKMMRRTELTVLEKGNFTFTDADIMLRD  
PFPRLYPDGDPOMACDRFGDPDSDNVMNGFTYVKSNNRSTIEFKFWNSRLDYPK  
MHODVFNQIKHKALVSEIGIQMRFPDYVYFGGFCQTSRBDINLWCTMHANCCVGLAKK  
LHDNLVLDDMRNYLSLSEPVKNTWTSVPKCA"

15812. 16087  
/gene="AT4g19970"  
intron /number=1  
16088. 16378  
/gene="AT4g19970"  
exon /number=2  
16379. 16756  
/gene="AT4g19970"  
intron /number=3  
16757. 17693  
/gene="AT4g19970"  
exon /number=4  
17694. 18160  
/gene="AT4g19970"

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-14 x ATCHRIV52 ..  
Align seg 1/1 to: ATCHRIV52 from: 1 to: 198427

34 GlyGIuLysArgGIuThrLysGIuLys 42  
|||||  
90028 GCGGMAAGCGAGACGACGAGAGAAA 90054

seq\_name: gb\_htg:AC073916  
seq\_documentation\_block:  
AC073916 206258 bp DNA linear HTG 29-JAN-2002  
LOCUS Homo sapiens chromosome 12q clone RP11-408118, WORKING DRAFT  
DEFINITION  
SEQUENCE 4 unordered pieces.  
ACCESSION  
AC073916  
VERSION  
AC073916.36 GI:18390150  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 206258)  
REFERENCE  
AUTHORS  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Aisburooks,S.L., Amaralunge,H.C., Are,u.R., Ayele,M., Banks,T.,  
Barbatoria,J., Benton,J., Bimage,K., Blankenbury,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carton,T.F., Carter,M., Cavazos,S.R., Checko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delanerie,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinin,H.H.,  
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,  
Homs,J.F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kraticovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lieu, C., Liu, J., Liu, W., Lonsdale, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Maxwell, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S., Oghu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rojfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tameris, A., Tameris, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanai, K., Vasquez, L., Vera, V., Villalon, D., Vinton, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished  
Direct Submission  
2 (bases 1 to 206258)  
Worley, K.C.

Direct Submission  
Submitted (05-JUL-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jan 29, 2002 this sequence version replaced gi:14717253.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

Project Information  
Center project name: HBW  
Center clone name: RP11-408118

Summary Statistics  
Sequencing vector: Plasmid: M77789  
Sequencing vector: M13: L08821  
Chemistry: Dye-terminator Big Dye: 834 of reads  
Assembly program: Phrap: version 0.990329Pfrst call to findPfrst

Consensus quality: 205705 bases at least Q40  
Consensus quality: 206814 bases at least Q30  
Consensus quality: 207486 bases at least Q20  
Estimated insert size: 207040; sum-of-coverage estimation  
Quality coverage: 0x in Q20 bases; agarose-IP estimation  
Quality coverage: 8.9x in Q20 bases; sum-of-coverage estimation

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 146278: contig of 146278 bp in length  
\* 146378: gap of unknown length  
\* 146379: contig of 3437 bp in length  
\* 181316: gap of unknown length  
\* 181416: contig of 21942 bp in length  
\* 203358: gap of unknown length  
\* 203458: contig of 2801 bp in length.

Location/Qualifiers  
1. 206258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12q"

BASE COUNT 43026 a 58439 c 59239 g 45228 t 306 others  
ORIGIN

alignment\_scores: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AC073916 ..  
Align seg 1/1 to: AC073916 from: 1 to: 206258

36 TysATGtGurhTyrGsluYsgluAan 44  
|||||  
139671 AACGCCAAMCCAGACAGACAGACAG 139697

seq\_name: gb\_in:AF008215

seq\_documentation\_block:  
LOCUS AF008215 2176 bp mRNA linear INV 21-JUN-1999  
DEFINITION Valimorpha necatrix mitochondrial Hsp70 homolog mRNA, complete  
CDs.  
ACCESSION AF008215 GI:2555076  
VERSION AF008215.1 GI:2555076  
KEYWORDS  
SOURCE  
ORGANISM  
Valimorpha necatrix  
Eukaryota; Microsporidia; Burenellidae; Valimorpha.  
REFERENCE  
AUTHORS  
TITLE  
Hirt, R.P., Healy, B., Vossbrinck, C.R., Canning, E.U., and Embley, T.M.  
A mitochondrial Hsp70 orthologue in Valimorpha necatrix: molecular evidence that microsporidia once contained mitochondria  
Curr. Biol. 7 (12), 995-998 (1997)

JOURNAL MEDLINE 98044288  
PUBMED 9382838  
REFERENCE 2 (bases 1 to 2176)  
AUTHORS Hirt, R.P. and Healy, B.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1997) Zoology, The Natural History Museum, Cromwell Rd, London SW7 5BD, UK

FEATURES  
source  
1. 2176  
/organism="Valimorpha necatrix"  
/db\_xref="taxon:6039"  
/tissue="spores"  
177..1910  
/codon\_start=1  
/product="mitochondrial Hsp70 homolog"  
/protein\_id="AAB81494.1"  
/translation="MTGKEISRRIGIDGTTNSCVSLMNNVPHIIEEGRTPTTS  
VFSKDKILVGDQAKOHLFPKNTIFASKRLIGRFEDLEKYLNLNLYDPTTHC  
NGDWAKRIDEKYSPAOIGAFILSKMNAEFLNSVVSAYVAVFDDMOQATK  
DAGTIGLELVNVEPTAAALVGLDSAGGIAYVDGGGTPDISLELDNGIFHV  
KSTNGNPIFGSGEDLNKLVDYINERKSKTGIDILKNNAYYRIKOEKIKRSLSTK  
LQSPINPIYVNDKSNYHKETITDEPERISREIINKTIDPCIKAMROAKTKNDI  
KCHLNGCATMPVYRKLYKRTGIRPSNTINPDAIAGACALOCAGVGVYDILL  
DVPISIGIEIGLVGKNTIIRNSYTPKRTGHSFSGSENGSKIDIPRYOGERMWKD  
MKYGAIAKIKNTPSAPACIKETIVTEADANGITKYVADQGISKKKQETIEIVSSGIN  
EEIKQVKEAKENDEKDSQKIALVYKLEIKTLERNIKERORLEETLKQGGFDI  
EEVANKTSKL 398 g 596 t

BASE COUNT 908 a 274 c 398 g 596 t  
ORIGIN

alignment\_scores: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000



seq\_name: gp\_in:010414

seq\_documentation\_block:

LOCUS 010414 28990 bp DNA linear INV 05-OCT-2001  
DEFINITION Caenorhabditis elegans cosmid F42A10, complete sequence.ACCESSION 010414  
VERSION 010414.1 GI:500733

KEYWORDS HTG.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderae; Caenorhabditis.

REFERENCE 1 (bases 1 to 28990)

AUTHORS The C. elegans Sequencing Consortium.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)JOURNAL MEDLINE  
99069613

REFERENCE 2 (bases 1 to 28990)

AUTHORS Latreille,P.  
TITLE The sequence of C. elegans cosmid F42A10  
JOURNAL Unpublished

REFERENCE 3 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Unpublished

REFERENCE 4 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL UnpublishedREFERENCE 5 (bases 1 to 28990)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-1994) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE 7 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUN-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 8 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (14-SEP-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 9 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 10 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 11 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 12 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 13 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 14 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 15 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 16 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

REFERENCE 17 (bases 1 to 28990)

AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USAThe 5' cosmid is Y4269A, 500 bp overlap; 3' cosmid is C23610, 200 bp  
overlap. Actual start of this cosmid is at base position 7 of  
CEL42A10; actual end is at 7862 of CELC23610.

## NOTES:

Coding sequences below are predicted from computer analysis, using  
the program GeneFinder(P. Green and L. Hillier, ms in preparation).

## NEIGHBORING COSMID INFORMATION

FEATURES  
source  
1. 28990  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/chromosome="III"  
/clone="F42A10"  
complement(699..3324)  
/gene="F42A10.3"  
complement(101669..836,1364..1441,2028..2514,  
3203..3324)  
/gene="F42A10.3"  
/note="exon 2 similar to small region of  
phenylethanolamine N-methyltransferase, B0303.4 and  
C1489.4; coded for by the following C. elegans cDNAs:  
cm1298, yk444b2.5, B1174689, B1175667"  
/codon\_start=1  
/product="Hypothetical protein F42A10.3"  
/protein\_id="AA19074.2"  
/db\_xref="GI:15617835"  
/translation="MSTLIDPTTSSDQAGTVLLPDDYKFKFDPDEYLOFYFSRDAI  
EDGTRVSLFALPFAQLMOTMRPSERELDIGAGTYVSALCPADVAKRVHLSDF  
DRNLDVLRKMRHRETDVPTIKVIRREGGPTPOVCIVTEKARGVSGGSHF  
ADVHMDVPELACKKVIDIVSIFLESACNRYEYCKVTMMHHLSSGSRVLSV  
LEDDEYNSGRKTFILNLESDQIIDLAVSLGVANAKVYVLDGEGVFMATKL"  
4374..4454  
/product="tRNA-Ser"  
4561..4645  
/product="tRNA-Ser"  
5475..5939  
/gene="F42A10.2"  
join(5475..5696,5827..6042,6389..6607,6684..6818,  
6866..7000,7076..7378,7422..7571,7673..7760,8233..8503,  
8615..8789,9889..9939)  
/gene="F42A10.2"  
/note="similar to ezrin (contains a domain found in band  
4.1, ezrin, moesin, radixin and talin); contains  
similarity to Pfam domain PF00373 (Band\_41), Score=234.3,  
E-value=4.1e-75, N=1; PF00769 (ERM), Score=489.2,  
E-value=1e-143, N=1; coded for by the following C. elegans  
cDNAs: yk34h5.3, yk34h5.5, yk52e5.5, yk214b9.3, yk214b9.5,  
yk406f4.5"  
/codon\_start=1  
/product="Hypothetical protein F42A10.2a"  
/protein\_id="AA19073.1"  
/db\_xref="GI:500735"  
/translation="MGSPSTFSRNTKPIYAKVSTNADLEKIVKWTGRHLPEAVC  
RILGIEPTVREGLOTPKKNIPCHLONDKTICGDDIKOTSGCTNPLIKKYPEDV  
EPETLIDARHRLFTQIRPAILSMNLCSPASVLAERAVAGDCTEGVETLID  
KLDPKSVTDQIDMSADMRIRKMSKACOSRESELESLKRAVODLMEGLTLP  
CNKNETDLHGTSAGGGLITKGVNRTTPPFSVSEKINQKRNKFKMVKVSTIS  
FRSETSIDSSLDLCIGTNHLLRROPDTLEVOQAQDAEDQORAAQKVALE  
RKQOEKREYKEMKOVYEWLELKAQENIRAEANDQALEARISEHLELMY  
QKSEVAEENRSLNNKSEALRMRKAREAITLAKQMSLADVSLANRKMVY  
SSSALILIDPNHQAQVSGHSGHPGSPGSGFHLQSLDTHHPQWMSLANKPSSA  
FSSPISNPQPLSHOQPOOMSQNTQAHQOQOVONQPPYVGGSCMIIIPDPSISQI  
FEQDTTLMLEKRSREYETRARIFKEHLELRGIDGLKRDQNVONGHREDAVHAQ  
NVAGEDKLETKRSMNGRTPROAQAQFGM"  
join(5475..5696,5827..6042,6389..6607,6684..6818,  
6866..7000,7076..7378,7422..7571,7673..7760,8233..8503,  
8615..8789,8995..9210)  
/gene="F42A10.2"

```

/note="coded for by the following C. elegans CDNAs:
yk06ff4.5, yk636e12.3, yk636e12.5"
/codon_start=1
/product="Hypoethetical protein F42A10.2b"
/protein_id="AK68385.1"
/db_xref="GI:14574194"
/note="MGF51F5RNTKRPYAKVSTMDADLEKYTEKTMTGRHLFEAVC
RIIGLEPTWYEGLOFTKRNKIPCMQLOMDKTDIGODLOKSDGTFNPFLVYKRPDY
EPILLDPTATPELFIOTKREILISMNLXCSPEAVALASPAVQAMHDCITFEVPIDLD
KHLPKSTIYDMSADMRDRIRKWRMSRNAGQREBELELYLVAQDLERYGLIYPI
CNNEEDLHGISAQGLGYKGNVNRITPRFESWSEIKVNIQFNKRFHMTKYDSTIS
FRNSETSIDSISDLDCIGTHNLVLRROPELVQOMRSQAKDKORAAEQAKVALE
RKSEQVEKERYLSMNKKOKVWEVWLELMACQENIKAEAEANQALAEKRAHSEFTLMYK
OKESEVEACNRLSKNNKSEBALRMEKRAEELIIEAKOMSLADYSIDANKMYQ
SSQSLATIDPNMHOOVYSGHSGHGMSTGSPHOLSDPTHRTPOOMSRSLMPSEA
FSSPSIPNPLSHOQPOOMQONTQAHQOQYVONONPPYTGSSMTIIPPDYSIOI
FEQDTTLMELKRSSETETRIARIFKEHLELRCDIDBLKRVGVNQGNEHDAVIAQ
NVAHGFRKFTPMRKSKRSASSIGSGSSDLSLENTLVARRHVAQRIPRVSTVSTSS
SILSHVMHLRHOOSYDLGNTGMEQAD"
/complement (10885, .14126)
/gene="F42A10.4"
complement (join(10885, .1149, 11747, .11891, 12037, .12060,
12115, .12501, 12745, .12850, 12904, .13960, 14067, .14126))
/gene="F42A10.4"
/note="coded for by the following C. elegans CDNAs:
yk3f11.3, yk3f11.5, yk154a4.5, yk154a4.3, yk154d5.5,
yk154d5.3, yk183b4.3, CEMS19F, yk324d1.3, yk408g12.5,
yk429g10.5, yk445b7.5, yk522f8.5, yk602g10.3, yk602g10.5,
yk681g2.3, yk681g2.5, yk728g6.5, yk735b12.5, yk735b12.3,
yk728g6.3"
/codon_start=1
/product="Hypoethetical protein F42A10.4"
/protein_id="AA19075.1"
/db_xref="GI:500737"
/note="MTIDTINESDNSTPNSPGLSEASRPSLSNASKMVRITTDYADVE
FISQNDVIEKPRMDPLHVRKLEMTWKKARARARTYVITDPMDSNHYVYAKRYR
YSALRQMTKEEDIVDYLHPDSFARGANREYRIKKSKEHTSDMNSNYAKRYICQV
DRYLEDVYLRQMDAKLMAEYRNIRNPKRITDLYQKCVIMIDYKSPSLYHLHTEG
KYIKYNSNGSVSNAARLYQARSHFTFEQSGHQMWVVDIQGQDLYTDPQITHTVGT
DYDGDNDIGTGMALEFHSRNCNDICEFMDLSNPELSPPELEATEVAMVAAKQKCI
VPTPVEFARRNRITSSSECVAHVEHGISMOLKRRKLNOSSTEDLSAKSHNECQPECIP
VPSOLCEPSEDEDEDEEDYVPTNITISLOQAAANLENDDEPVQYTGHOEVLQGIH
IDSLSRYHELGRPEVDESEHEKEMLEGSNDNRAPVIRKDKOSAIFLIDARCGILLEVL
TSARYHELGRPELLEKTYVTDLTPNGSGEENIRADKQKCPDLEDSQDLMEITAE
MGCKGAMLYMAHAHYETGOHLGPNRRDIKYSITDWTQRVGFOEEELSDSCGKTTSS
FAPLITREILAKAEMTKEGSYGLNODFERAYGLFNEAEALAEAMNGKLANKYERA
EMQEG"
complement (14143, .14144)
/note="S1L trans-splice site; see ESNs yk681g2.5,
yk735b12.5 and yk728g6.5"
/gene="F42A10.5"
complement (17398, .18367)
/gene="F42A10.5"
/gene="F42A10.5"
complement (join(17398, .17529, 17631, .18093, 18144, .18367)
/note="coded for by the following C. elegans CDNAs:
yk44f7.3, yk74f7.5, yk112g6.3, yk112g6.5, yk158h11.3,
yk158h11.5, yk370c4.3, yk370c4.5, yk453c7.3, yk513d12.5,

```

LOCUS	CE11C1	40852 bp	DNA	linear	INV 23-JAN-2002
DEFINITION	Caenorhabditis elegans cosmid F11C1, complete sequence.				
ACCESSION	254270				
VERSION	254270.1	GI:3642282			
KEYWORDS	HTG; Dentrification; Membrane glycoprotein; Nlr; Nuclear hormone receptor.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans.				
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.				
AUTHORS	1 (sites)				
TITLE	none.				
JOURNAL	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium				
MEDLINE	Investigating Biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)				
REMARK	99069613				
REFERENCE	The C.elegans Sequencing Consortium.				
AUTHORS	2 (bases 1 to 40852)				
TITLE	Palmer,S				
JOURNAL	Direct Submission				
COMMENT	Submitted (29-SEP-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jtesanger.ac.uk or twenematode.wustl.edu On Sep 22, 1998 this sequence version replaced gi:1001858. Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. This sequence is the entire insert of clone F11C1. The start of this sequence (1..104) overlaps with the end of sequence AL023817. The end of this sequence (40749..40852) overlaps with the start of sequence AL008868. For a graphical representation of this sequence and its analysts see: <a href="http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F11C1">http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F11C1</a> IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. Location/Qualifiers 1..40852 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="X" /clone="F11C1" join(980..1235,1367..1611) /gene="F11C1.4" join(980..1235,1367..1611) /gene="F11C1.4" /codon_start=1 /protein_id="CAA91029.1" /db_xref="GI:3875725" /db_xref="SPTREMBL:Q19347" /translation="MTISPLFALVLPPLHIKFTYHVGNGDGYTFDGRRLIN LLAERLGRTRVOEFYVLSHINHLVLTIDPTSCSESHKVAVERSLADVOYHKIDFPYK EYPRGNRYHYMGHGGGAYMLSLIPYIKDPSLNKKAACLPETIERMTESNHGIRLKK IVSVRR" join(4336..4463,5151..5312,5860..5935,5980..6062,6249..6327,6386..6513,6562..6717,6800..6929,7378..7692,7743..7816,7965..8229) /gene="F11C1.3" join(4336..4463,5151..5312,5860..5935,5980..6062,6249..6327,6386..6513,6562..6717,6800..6929,7378..7692,7743..7816,7965..8229) /gene="F11C1.3" /note="contains similarity to Pfam domain: PF01130 (CD36				



seq\_documentation\_block: 50962 bp DNA linear HTG 16-OCT-2001  
 LOCUS AC090028 Homo sapiens chromosome 3q clone RP11-29962, \*\*\* SEQUENCING IN  
 DEFINITION  
 AC090028 AC090028.2 GI:14190614  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 50962)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alstbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,J., Bimge,K., Blankenburg,K., Bonin,D., Bouck,J.,  
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Den,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
 Hollins,B., Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W.,  
 Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
 Ogun,M., Okunolu,G., Oragunye,N., Oyedro,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., RojudoKan,I., Ruiz,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,  
 Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Viallon,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 50962)  
 Worley,K.C.  
 Direct Submission  
 Submitted (10-FEB-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 23, 2001 this sequence version replaced g1:12739654.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HDEF  
 Center clone name: RP11-29962  
 ----- Summary Statistics  
 Sequencing vector: M13, L08821  
 Chemistry: Dye-terminator Big Dye: 99% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 27482 bases at least Q40

Consensus quality: 35893 bases at least Q30  
 Consensus quality: 39768 bases at least Q20  
 Estimated insert size: 18665; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agorose-fp estimation  
 Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 48 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 647: contig of 647 bp in length  
 648 747: gap of unknown length  
 748 1436: contig of 689 bp in length  
 1437 1536: gap of unknown length  
 1537 2170: contig of 633 bp in length  
 2170 2268: gap of unknown length  
 2268 2928: contig of 659 bp in length  
 2928 3028: gap of unknown length  
 3028 3666: contig of 638 bp in length  
 3666 3767: gap of unknown length  
 3767 4441: contig of 675 bp in length  
 4441 5224: gap of unknown length  
 5224 5324: gap of unknown length  
 5324 5953: contig of 629 bp in length  
 5953 6053: gap of unknown length  
 6053 6695: contig of 642 bp in length  
 6695 6796: gap of unknown length  
 6796 7431: contig of 636 bp in length  
 7431 7531: gap of unknown length  
 7531 8198: contig of 667 bp in length  
 8198 8298: gap of unknown length  
 8298 8949: contig of 651 bp in length  
 8949 9049: gap of unknown length  
 9049 9692: contig of 643 bp in length  
 9692 9792: gap of unknown length  
 9792 10508: contig of 716 bp in length  
 10508 11694: gap of unknown length  
 11694 11794: contig of 1086 bp in length  
 11794 12537: gap of unknown length  
 12537 12637: gap of unknown length  
 12637 13282: contig of 645 bp in length  
 13282 13382: gap of unknown length  
 13382 14186: contig of 804 bp in length  
 14186 14286: gap of unknown length  
 14286 15068: contig of 782 bp in length  
 15068 15168: gap of unknown length  
 15168 16268: contig of 1100 bp in length  
 16268 17150: gap of unknown length  
 17150 17250: contig of 782 bp in length  
 17250 18072: gap of unknown length  
 18072 18172: contig of 822 bp in length  
 18172 18868: gap of unknown length  
 18868 19868: gap of 696 bp in length  
 19868 19968: contig of 900 bp in length  
 19968 20865: gap of unknown length  
 20865 20965: contig of 897 bp in length  
 20965 21680: gap of unknown length  
 21680 21780: gap of 715 bp in length  
 21780 22809: gap of unknown length  
 22809 22910: contig of 1029 bp in length  
 22910 24084: gap of unknown length  
 24084 24184: contig of 1175 bp in length  
 24184 25112: gap of unknown length  
 25112 25112: contig of 928 bp in length



\* 25113 25212: gap of unknown length  
\* 25213 26004: contig of 792 bp in length  
\* 26005 26104: gap of unknown length  
\* 26105 26945: contig of 841 bp in length  
\* 26946 27045: gap of unknown length  
\* 27046 28248: contig of 1203 bp in length  
\* 28249 28348: gap of unknown length  
\* 28349 29254: contig of 906 bp in length  
\* 29255 29355: gap of unknown length  
\* 29356 30383: contig of 1028 bp in length  
\* 30383 30482: gap of unknown length  
\* 30483 31401: contig of 918 bp in length  
\* 31401 31501: gap of unknown length  
\* 31501 32882: contig of 1382 bp in length  
\* 32883 32982: gap of unknown length  
\* 32983 34198: contig of 1216 bp in length  
\* 34199 34298: gap of unknown length  
\* 34299 35713: contig of 1415 bp in length  
\* 35714 35813: gap of unknown length  
\* 35814 36984: contig of 1171 bp in length  
\* 36985 37084: gap of unknown length  
\* 37085 38284: contig of 1200 bp in length  
\* 38285 38384: gap of unknown length  
\* 38385 40364: gap of unknown length  
\* 40365 40464: gap of unknown length  
\* 40465 41816: contig of 1352 bp in length  
\* 41817 41916: gap of unknown length  
\* 41917 42742: contig of 826 bp in length  
\* 42743 42842: gap of unknown length  
\* 42843 44118: contig of 1276 bp in length  
\* 44119 44218: gap of unknown length  
\* 44219 45927: contig of 1709 bp in length  
\* 45928 46027: gap of unknown length  
\* 46028 47154: contig of 1127 bp in length  
\* 47155 47254: gap of unknown length  
\* 47255 48939: contig of 1685 bp in length  
\* 48940 49039: gap of unknown length  
\* 49040 50962: contig of 1923 bp in length.  
Location/Qualifiers  
1. 50962  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3q"  
/clone="RP11-299G2"

BASE COUNT 13625 a 9143 c 9047 g 14381 t 4766 others  
ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AC090028/rev ..  
Align seq 1/1 to reverse of: AC090028 from: 1 to: 50962

36 LysArgGluThrLysGluLysGlu 43  
|||||||  
29713 AAGGAGAAACAAAGAGAAAGAA 29690

seq\_name: gb\_pr:HSDJ223E3

seq\_documentation\_block:  
LOCUS HSDJ223E3 59621 bp DNA linear PRI 04-APR-2001  
DEFINITION Human DNA sequence from clone RP1-223E3 on chromosome 6p13-15  
Contains the gene for a putative secreted protein ZSIC13, ESTs and  
GSSs, complete sequence.  
ACCESSION AL121939  
VERSION AL121939.12 GI:10862831  
KEYWORDS HTG.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 59621)  
REFERENCE  
AUTHORS Dunn, M.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Oct 17, 2000 this sequence version replaced gi:10732514.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Mp: MOPREP; Information  
on the MOPREP database can be found at  
http://www.sanger.ac.uk/Projects/C.elegans/MOPREP This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/MOP/Ch6  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP1-223E3 is from the  
library RP1-1 constructed by the group of Pieter de Jong. For  
further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RP1-223E3. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP1-223E3 is at 1 in this sequence. The  
true left end of clone RP1-120N9 is at 59522 in this sequence.  
Location/Qualifiers  
1. 59621  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="q15-15"  
/clone="RP1-223E3"  
/clone\_lib="RP1-1"  
2. 168  
/note="AluB repeat: matches 107. .298 of consensus"  
172. .879  
/note="AluB repeat: matches 4490. .5213 of consensus"  
880. .1098  
/note="MER30 repeat: matches 4. .230 of consensus"  
1099. .1156  
/note="LIM4 repeat: matches 5213. .5269 of consensus"  
1329. .1552  
/note="AluSd/x repeat: matches 83. .310 of consensus"  
1638. .1746  
/note="MER81 repeat: matches 2. .109 of consensus"  
1764. .2202  
/note="LIM5 repeat: matches 7306. .7740 of consensus"  
2400. .2830  
/note="LIPa10 repeat: matches 5730. .6158 of consensus"  
3131. .3443  
/note="AluSx repeat: matches 1. .312 of consensus"  
3671. .3806  
/note="LIM3 repeat: matches 7761. .7907 of consensus"  
4513. .4741  
/note="AluB repeat: matches 84. .303 of consensus"  
4872. .5037  
/note="AluB repeat: matches 126. .289 of consensus"

repeat\_region 5046..5435  
/note="LMC2 repeat: matches 5910..6324 of consensus"  
repeat\_region 5439..5813  
/note="MLT1B repeat: matches 1..390 of consensus"  
repeat\_region 5828..6022  
/note="MST-INTERNAL repeat: matches 1443..1637 of consensus"  
repeat\_region 6023..6314  
/note="AluX repeat: matches 1..293 of consensus"  
repeat\_region 6311..6391  
/note="MLT1-INTERNAL repeat: matches 1046..1111 of consensus"  
repeat\_region 6392..6685  
/note="AluSx repeat: matches 1..290 of consensus"  
repeat\_region 6686..6788  
/note="MLT1-INTERNAL repeat: matches 961..1046 of consensus"  
repeat\_region 6790..7091  
/note="AluSp repeat: matches 1..299 of consensus"  
repeat\_region 7539..8133  
/note="MLT1B-internal repeat: matches 1..728 of consensus"  
repeat\_region 8168..8363  
/note="MLT1B repeat: matches 194..390 of consensus"  
repeat\_region 8427..8451  
/note="MLT1B repeat: matches 168..188 of consensus"  
repeat\_region 8452..8758  
/note="AluJo repeat: matches 1..301 of consensus"  
repeat\_region 8759..8910  
/note="MLT1B repeat: matches 1..168 of consensus"  
repeat\_region 8911..8998  
/note="LMC2 repeat: matches 5826..5911 of consensus"  
repeat\_region 8999..9299  
/note="AluSx repeat: matches 1..300 of consensus"  
repeat\_region 9300..9949  
/note="LMC2 repeat: matches 5189..5826 of consensus"  
repeat\_region 9950..10223  
/note="AluJo repeat: matches 1..298 of consensus"  
repeat\_region 10224..10633  
/note="LMC2 repeat: matches 4787..5189 of consensus"  
repeat\_region 10634..10920  
/note="AluSx repeat: matches 12..296 of consensus"  
repeat\_region 10921..11272  
/note="LMC2 repeat: matches 4362..4787 of consensus"  
repeat\_region 11292..11412  
/note="FLAM-A repeat: matches 13..128 of consensus"  
repeat\_region 11413..12017  
/note="LMC2 repeat: matches 3787..4359 of consensus"  
repeat\_region 12046..12343  
/note="AluSg1 repeat: matches 1..294 of consensus"  
repeat\_region 12582..18699  
/note="LIPAS repeat: matches 11..6143 of consensus"  
repeat\_region 18731..19020  
/note="AluJb repeat: matches 1..293 of consensus"  
misc\_feature 18754..19116  
/note="match: GSS: Em:AQ521768"  
repeat\_region 19702..20069  
/note="THE1B repeat: matches 1..364 of consensus"  
repeat\_region 20223..20631  
/note="MLT2FB repeat: matches 1..410 of consensus"  
repeat\_region 21092..21183  
/note="MIR repeat: matches 23..112 of consensus"  
repeat\_region 21243..21463  
/note="MER20 repeat: matches 2..195 of consensus"  
repeat\_region 22606..22805  
/note="MIR repeat: matches 49..250 of consensus"  
repeat\_region 22909..23196  
/note="AluJo repeat: matches 1..294 of consensus"  
repeat\_region 23505..23562  
/note="29 copies 2 mer at 70% conserved"  
repeat\_region 23700..23827  
/note="LIME repeat: matches 5678..5804 of consensus"  
repeat\_region 23841..24023  
/note="L2 repeat: matches 1904..2095 of consensus"

repeat\_region 24642..24946  
/note="AluSx repeat: matches 1..300 of consensus"  
repeat\_region 25088..25297  
/note="MIR repeat: matches 32..247 of consensus"  
repeat\_region 27241..27530  
/note="AluJo repeat: matches 1..288 of consensus"  
repeat\_region 28031..28100  
/note="L2 repeat: matches 2625..2693 of consensus"  
repeat\_region 28236..28482  
/note="MER8 repeat: matches 4..238 of consensus"  
repeat\_region 28635..28670  
/note="18 copies 2 mer tt 80% conserved"  
repeat\_region 29295..29412  
/note="59 copies 2 mer cc 65% conserved"  
misc\_feature complement(29374..29836)  
/note="match: GSS: Em:AQ493139"  
repeat\_region 30071..30203  
/note="LMC5 repeat: matches 7762..7893 of consensus"  
misc\_feature 31005..31642  
/note="match: GSS: Em:B66029"  
misc\_feature 31005..31469  
/note="match: GSS: Em:AQ232280"  
repeat\_region 32407..32692  
/note="AluJb repeat: matches 1..300 of consensus"  
repeat\_region 32729..33022  
/note="AluSx repeat: matches 4..300 of consensus"  
repeat\_region 33097..33203  
/note="MLT1D repeat: matches 396..505 of consensus"  
repeat\_region 33204..33510  
/note="AluY repeat: matches 1..305 of consensus"  
repeat\_region 33511..33886  
/note="MLT1D repeat: matches 40..396 of consensus"  
repeat\_region 33990..34197  
/note="MIR repeat: matches 13..252 of consensus"  
repeat\_region 35679..35938  
/note="MIR repeat: matches 8..262 of consensus"  
repeat\_region 36923..37001  
/note="AluSc repeat: matches 223..299 of consensus"  
repeat\_region 37212..37407  
/note="MER5A repeat: matches 1..188 of consensus"  
repeat\_region 38378..38435  
/note="29 copies 2 mer cc 70% conserved"  
repeat\_region 38441..38712  
/note="AluJo repeat: matches 1..280 of consensus"  
misc\_feature complement(38574..39022)  
/note="match: GSS: Em:AQ572560"

alignment\_scores:                   Quality: 8.00                   Length: 8  
                                  Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000   Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x HSDJ223E3/rev ..

Align seg 1/1 to reverse of: HSDJ223E3 from: 1 to: 59621


36 LysArgGluThrLysGluLysGlu 43  
|||||  
54121 AAGAGAGAAACAAAGAGAGAA 54098

Mon Aug 5 11:51:44 2002

us-08-973-363-14.olip2n.rge

---

Page 17





OM of: US-08-973-363-14 to: N\_Geneseq\_032802.\* out\_format : pfs

Date: Aug 3, 2002 7:18 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

-MODEL=itame+.p2n.model -DEV=x1h  
-O=/cgn2.1/USPTO.spool/US08973363/runat\_01082002\_080123\_19849/app.query.fasta\_1.638  
-DB=N\_Geneseq\_032802 -OFMT=fastap -SUFFIX=olip2n.rng  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000  
-LOAPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000  
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000  
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORMEXT -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US08973363 @CGNL 1 186 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMOUT=120 -WARN\_TIMOUT=30 -NO\_XLPEXT -WAIT -THREADS=1

#### Search information block:

Query: US-08-973-363-14  
Query length: 43  
Database: N\_Geneseq\_032802.\*  
Database sequences: 1738436  
Database length: 858457221  
Search time (sec): 521.050000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

#### Score list:

Sequence	Strd Orig	ZScore	EScore	len	1	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT2753 +			33.00	599.83	4.6e-25	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT2752 +			11.00	205.43	0.0043	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS1339 +			8.00	143.39	12.24	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS1339 +			8.00	135.17	35.09	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAS2842 +			7.00	130.46	64.20	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT70025 +			7.00	130.04	67.72	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS39061 -			7.00	129.24	75.10	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAT0282 +			7.00	127.85	89.79	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAC0952 +			7.00	127.20	97.51	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAC0953 +			7.00	126.17	111.27	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAC0953 +			7.00	124.43	139.14	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAC2409 +			7.00	124.42	139.32	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAT1881 -			7.00	124.28	141.83	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT7229 -			7.00	124.17	143.81	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT7229 -			7.00	124.10	145.24	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAC5213 -			7.00	123.94	148.10	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT1207 +			7.00	123.33	160.20	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH03878 +			7.00	123.29	161.05	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH03878 +			7.00	122.78	172.07	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAH07228 -			7.00	121.88	192.97	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAX0159 -			7.00	121.78	195.57	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAS9160 +			7.00	120.69	224.78	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAX4976 -			7.00	119.66	256.66	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAT86702 -			7.00	119.39	265.66	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT1843 -			7.00	119.38	265.66	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS1393 +			7.00	119.09	279.66	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAS1176 +			7.00	118.94	281.50	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT4709 -			7.00	118.58	294.77	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH14022 +			7.00	118.43	300.55	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH13715 +			7.00	117.59	334.67	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS2267 +			7.00	117.01	360.13	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS3251 +			7.00	116.81	366.33	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH7226 -			7.00	116.56	381.93	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT7225 -			7.00	116.19	395.56	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT7225 -			7.00	115.96	400.52	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT7472 +			7.00	115.94	412.33	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT7472 +			7.00	115.73	424.84	

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS1235 +	7.00	115.37	444
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAS3499 +	7.00	115.16	456
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH27507 +	7.00	115.12	459
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH17087 +	7.00	115.03	464
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH09681 -	7.00	113.10	594
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAS2952 -	7.00	110.92	787
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAH27506 +	7.00	110.73	806

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT2753

seq\_documentation\_block:

ID AAT42753 standard; CDNA: 137 BP.

AC AAT42753;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-W clone CC14 3' motif.

KN Bird: sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KN CHD-W; chromodomain-Helicase-DNA binding on the W chromosome, ss.

OS Gallus sp.

Key Location/Qualifiers

FT misc\_difference 52

FT /\*tag= a

FT /note= "base 52 disrupts the reading frame for

FT the translated amino acid sequence given

FT in Fig 7"

PN MO639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96MO-GB01341.

PR 06-JUN-1995; 96GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tlwari B;

XX WP1: 1997-043127/04.

XX P-PSDB: AAW08145.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

XX birds - used for sex determ. and to control sex of progeny

XX Disclosure: Fig 7; 76pp; English.

XX A composite (incomplete) nucleotide sequence (AAT42753) and putative

XX translation (AAW08145) sequence are provided of a motif that is found

XX spliced to the 3' end of chicken CHD-W clone CC14 (see also AAT42754).

XX There are no splice donor or acceptor sites within the motif

XX suggesting it is a final rather than an intermediary product of

XX splicing. The motif is also found as an insert in some CHD-1A clones

XX (see also AAT42752).

XX Sequence 137 BP; 75 A; 10 C; 34 G; 18 T; 0 other;

XX alignment\_scores: 32.00 Length: 46

XX Quality: 0.711 Ratio: 1

XX Percent Similarity: 97.826 Percent Identity: 97.826

XX alignment\_block:

XX US-08-973-363-14 x AAT42753 ..

XX Align seg 1/1 to: AAT42753 from: 1 to: 137

XX 1 AapcylleValserVallyleHnIspOnHlsLysIleLysAlaLlu.L 17

```
|||||
1 GATGGATGTTTCAGTGAACATCCACATATAAAATAAGACGAGAAA 50
17 YSGIUAASGLUGIULysAspGIupProGIuileGIyLeLysLysGLuAa 33
51 AAGAAATGTAAGAAAAAGATGATGACGAGATTGGTATTAAGAGAACT 100
34 GYGIULysArgGIuThrLysGLuLysGLuAsnLys 45
101 GGAGAAAAAGAGACAAAGAGAGAAATAAAG 136

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA1997.DAT:AAAT42752

seq_documentation_block:
ID AAT42752 standard; cDNA; 265 BP.
XX
XX AAT42752;
XX
XX 12-MAR-1997 (first entry)
XX
XX Chicken CHD-1A insert motif.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
XX Gallus sp.
XX
XX Key Location/Qualifiers
XX misc_difference 52 /*tag- a
XX FT /note- "base 52 disrupts the reading frame for
XX FT the translated amino acid sequence given
XX FT in Fig 7"
XX
XX PN WO9639505-A1.
XX
XX PD 12-DEC-1996.
XX
XX PF 05-JUN-1996; 96WO-GB01341.
XX
XX PR 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX P-PSDB; AAW08144.
XX
XX DR Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX PT Disclosure: Fig 7; 76pp; English.
XX
XX PS A composite nucleotide sequence (AAAT42752) and putative translation
XX (AAW08144) sequence are provided of a motif that is found spliced to
XX a proportion of chicken CHD-1A clones. The motif is inserted
XX between bases 4327 and 4328 of the CHD-1A composite sequence
XX (AAAT42751). None of the 7 CHD-1 clones examined contained the
XX complete motif. There are no splice donor or acceptor sites within
XX the motif suggesting it is a final rather than an intermediary
XX product of splicing. The motif is also found at the 3' end of the
XX CC CHD-W clone CC14 (see also AAT42753).
XX
XX Sequence 265 BP; 158 A; 16 C; 61 G; 30 T; 0 other;
```

## alignment\_scores:

```
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

## alignment\_block:

```
US-08-973-363-14 x AAT42752 ..
Align seg 1/1 to: AAT42752 from: 1 to: 265
35 GIUysArgGIuThrLysGLuLysGLuAsnLys 45
|||||
104 GAAAAAGAGAGACAAAGAGAGAAATAAAG 136

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embL/NA2001A.DAT:AAS31339

seq_documentation_block:
ID AAS31339 standard; cDNA; 657 BP.
XX
XX AAS31339;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human cDNA encoding a novel extracellular matrix protein, Seg ID No 153.
XX
XX Human; secreted extracellular matrix protein; ss; Immunomodulatory;
XX Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
XX cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
XX antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
XX human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
XX cancers; hyperproliferative disorder; breast neoplasm; melanoma;
XX Sezary syndrome; Gaucher's disease; neurological diseases;
XX Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
XX cardiac arrest; tachycardia; angina; infection; corneal infections;
XX wound healing; immunogen; gene therapy; antisense; food additive.
XX
XX OS Homo sapiens.
XX
XX PN WO200155368-A1.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01348.
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226686.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227009.
```



```
Align seg 1/1 to: AAS1339 from: 1 to: 657
      35 GlutylsATGtUThrLyGLuLys 42
      ||||||||||||||||||
496 GAAAAAGCAACGAAAGAGAGAA 519

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-nembl/NA2001A.DAT.AAK89386
seq_documentation_block:
ID   AAK89386 standard; DNA; 2121 BP.
XX
AC   AAK89386;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2962.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   MO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001MO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205513.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226688.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228824.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
```

```
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   27-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241121.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   01-NOV-2000; 2000US-0244674.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
```



```
PR 17-NOV-2000; 2000US-0248244.
PR 17-NOV-2000; 2000US-0248245.
PR 17-NOV-2000; 2000US-0248264.
PR 17-NOV-2000; 2000US-0248265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM:
XX
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosis, treating, preventing and/or prognostizing disorders of the
XX digestive system, particularly cancer and cancer metastases -
XX
XX Disclosure: SEQ ID NO 2962; 986bp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic
XX infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX ulcerative colitis. The present sequence is a genomic DNA fragment
XX encoding a digestive system antigen of the invention.
XX
XX Sequence 2121 BP; 906 A; 306 C; 407 G; 502 T; 0 other;
SQ
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-14 x AAK89386 ..
Align seg 1/1 to: AAK89386 from: 1 to: 2121
35 Gtutygargltutrltysgtllys 42
|||||
1368 GAAAAAAGGAAACGAAAGAGAAA 1391
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAK25842
seq_documentation_block:
ID AAK25842 standard; cDNA; 294 BP.
XX
XX AAC25842;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 29917.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
```

```
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J:
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 29917; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 294 BP; 96 A; 51 C; 72 G; 75 T; 0 other;
SQ
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-14 x AAC25842 ..
Align seg 1/1 to: AAC25842 from: 1 to: 294
29 lleytlysgluagllygla 35
|||||
170 ATAAAGAAAGACGCTGTGAA 190
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK70025
seq_documentation_block:
ID AAK70025 standard; DNA; 312 BP.
XX
XX AAK70025;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24837.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
```

PE 17-JAN-2001; 2001MO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251038.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251656.  
PR 08-DEC-2000; 2000US-0251668.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC;  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI, 2001-483426/52.  
DR  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT

PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
PS Disclosure: SEQ ID NO 24837; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I), by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 312 BP; 133 A; 40 C; 82 G; 57 T; 0 other;  
  
alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-14 x AAK70025 ..  
  
Align seg 1/1 to: AAK70025 from: 1 to: 312  
  
37 ArgGluThrLysGluLysGlu 43  
|||||  
50 AGGAGAGCCAGCAGAGAGAGAG 70  
  
seq\_name: /STDS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAK39061  
seq\_documentation\_block:  
ID AAK39061 standard; cDNA; 350 BP.  
XX  
AC AAK39061;  
XX  
XX 17-DEC-2001 (first entry)  
XX  
DE Novel human diagnostic and therapeutic gene #2119.  
XX  
XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO20016753-A2.  
XX  
XX 13-SEP-2001.  
XX  
XX 09-MAR-2001; 2001WO-US07787.  
XX  
XX 09-MAR-2000; 2000US-0188609.  
XX  
XX (CHIR) CHIRON CORP.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
XX Rainhardt C, Randazzo E, Kennedy GC, Pol D, Kassam A, Lamson G;  
XX Dimech R, Crkvenjakov R, Dickson M, Dimech S, Labat I;  
XX Lesnikowicz D, Kita D, Garcia V, Jones WL, Stache-Crain B;  
XX WPI; 2001-530177/58.  
XX

PT New polynucleotides and polypeptides, useful for diagnosis and  
PT treatment of breast, lung and colon cancer -  
PS Claim 1; Page 1126; 1193pp; English.  
XX  
XX  
CC The invention relates to new polynucleotides and polypeptides, useful for  
CC diagnosis and treatment of breast, lung and colon cancer. The sequences  
CC can be used in detecting differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample derived from a  
CC cell suspected of being cancerous. They can also be used to inhibit  
CC tumor growth by modulating expression of a gene product. AAK36943-  
CC AAK39338 represent novel human diagnostic and therapeutic coding  
CC sequences of the invention.  
XX  
SQ Sequence 350 BP; 53 A; 84 C; 60 G; 153 T; 0 other;  
  
alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-14 x AAK39061/rev ..  
  
Align seg 1/1 to reverse of: AAK39061 from: 1 to: 350  
  
31 LysGluAlaGlyGluLysArg 37  
|||||  
260 AAGAGAGCCAGCAGAGAGAGA 240  
  
seq\_name: /STDS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAF10282  
seq\_documentation\_block:  
ID AAF10282 standard; cDNA; 427 BP.  
XX  
XX AAF10282;  
XX  
XX 13-MAR-2001 (first entry)  
XX  
XX  
DE Fusarium venenatum EST SEQ ID NO:2805.  
XX  
XX  
XX Multiple gene expression; filamentous fungal cell; EST.  
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
XX Aspergillus oryzae; trichoderma reesei; identification; recombination;  
XX culture condition; environmental stress; spore morphogenesis;  
XX metabolic pathway engineering; catalytic pathway engineering; ss.  
XX  
XX Fusarium venenatum.  
XX  
XX WO200056762-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07781.  
XX  
XX 22-MAR-1999; 99US-0273623.  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
XX (NOVO ) NOVO NORDISK AS.  
XX  
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
XX WPI; 2000-594572/56.  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
XX uses fluorescence-labeled nucleic acids isolated from the cells and a  
XX substrate of expressed sequence tags -  
XX Claim 86; Page 1400; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential  
XX



XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A)<sup>+</sup> RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC cDNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

SQ Sequence 542 BP; 151 A; 105 C; 130 G; 155 T; 1 other;

Alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment\_block:  
US-08-973-363-14 x AAC00953 ..

Align seg 1/1 to: AAC00953 from: 1 to: 542

13 TlclysAlaGluTysGluAsn 19  
|||||  
431 ATCAGGCTGAAAAAGAAAT 451

seq\_name: /SIDSI/gcdata/hold-geneseq/genesequ-emb1/NA2000.DAT:AAC54696

seq\_documentation\_block:  
ID AAC54696 standard; DNA; 695 BP.

XX AAC54696;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78761.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW Protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
PD  
XX 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128645.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139829.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 02-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 22-JUL-1999; 99US-0145218.  
PR 22-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 16-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 7.00  
Ratio: 1.000  
Percent Similarity: 100.000  
Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAC54696 ..

Align seg 1/1 to: AAC54696 from: 1 to: 695

26 GUILLEGlyIleLysGlu 32  
|||||  
190 GAGATTGGATCAAAAGAG 210

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC52409

seq\_documentation\_block:  
ID AAC52409 standard; DNA: 696 BP.  
XX  
AC AAC52409;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 71249.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999;	99US-0132485	PR 22-JUL-1999;	99US-0145087
PR 06-MAY-1999;	99US-0132486	PR 22-JUL-1999;	99US-0145089
PR 07-MAY-1999;	99US-0132487	PR 23-JUL-1999;	99US-0145192
PR 11-MAY-1999;	99US-0132488	PR 23-JUL-1999;	99US-0145145
PR 14-MAY-1999;	99US-0134256	PR 23-JUL-1999;	99US-0145218
PR 14-MAY-1999;	99US-0134219	PR 26-JUL-1999;	99US-0145524
PR 14-MAY-1999;	99US-0134219	PR 26-JUL-1999;	99US-0145526
PR 14-MAY-1999;	99US-0134221	PR 27-JUL-1999;	99US-0145513
PR 14-MAY-1999;	99US-0134370	PR 27-JUL-1999;	99US-0145518
PR 18-MAY-1999;	99US-0134768	PR 27-JUL-1999;	99US-0145519
PR 19-MAY-1999;	99US-0134941	PR 28-JUL-1999;	99US-0145551
PR 20-MAY-1999;	99US-0135124	PR 02-AUG-1999;	99US-0146386
PR 21-MAY-1999;	99US-0135353	PR 02-AUG-1999;	99US-0146388
PR 24-MAY-1999;	99US-0135629	PR 02-AUG-1999;	99US-0146389
PR 25-MAY-1999;	99US-0136021	PR 03-AUG-1999;	99US-0147038
PR 27-MAY-1999;	99US-0136392	PR 04-AUG-1999;	99US-0147204
PR 28-MAY-1999;	99US-0136782	PR 04-AUG-1999;	99US-0147302
PR 01-JUN-1999;	99US-0137222	PR 05-AUG-1999;	99US-0147192
PR 03-JUN-1999;	99US-0137528	PR 05-AUG-1999;	99US-0147260
PR 04-JUN-1999;	99US-0137502	PR 06-AUG-1999;	99US-0147303
PR 07-JUN-1999;	99US-0137724	PR 06-AUG-1999;	99US-0147416
PR 08-JUN-1999;	99US-0138094	PR 09-AUG-1999;	99US-0147493
PR 10-JUN-1999;	99US-0138540	PR 09-AUG-1999;	99US-0147933
PR 10-JUN-1999;	99US-0138847	PR 10-AUG-1999;	99US-0148171
PR 14-JUN-1999;	99US-0139119	PR 11-AUG-1999;	99US-0148319
PR 16-JUN-1999;	99US-0139452	PR 12-AUG-1999;	99US-0148341
PR 16-JUN-1999;	99US-0139453	PR 13-AUG-1999;	99US-0148565
PR 17-JUN-1999;	99US-0139452	PR 13-AUG-1999;	99US-0148684
PR 18-JUN-1999;	99US-0139454	PR 16-AUG-1999;	99US-0149368
PR 18-JUN-1999;	99US-0139455	PR 17-AUG-1999;	99US-0149175
PR 18-JUN-1999;	99US-0139456	PR 18-AUG-1999;	99US-0149426
PR 18-JUN-1999;	99US-0139457	PR 20-AUG-1999;	99US-0149722
PR 18-JUN-1999;	99US-0139458	PR 20-AUG-1999;	99US-0149723
PR 18-JUN-1999;	99US-0139459	PR 20-AUG-1999;	99US-0149929
PR 18-JUN-1999;	99US-0139460	PR 23-AUG-1999;	99US-0149902
PR 18-JUN-1999;	99US-0139461	PR 23-AUG-1999;	99US-0149930
PR 18-JUN-1999;	99US-0139462	PR 25-AUG-1999;	99US-0150566
PR 18-JUN-1999;	99US-0139463	PR 25-AUG-1999;	99US-0150884
PR 18-JUN-1999;	99US-0139750	PR 26-AUG-1999;	99US-0151065
PR 18-JUN-1999;	99US-0139763	PR 27-AUG-1999;	99US-0151066
PR 21-JUN-1999;	99US-0139817	PR 27-AUG-1999;	99US-0151080
PR 22-JUN-1999;	99US-0139899	PR 30-AUG-1999;	99US-0151303
PR 23-JUN-1999;	99US-0140353	PR 31-AUG-1999;	99US-0151438
PR 23-JUN-1999;	99US-0140354	PR 01-SEP-1999;	99US-0151930
PR 24-JUN-1999;	99US-0140693	PR 07-SEP-1999;	99US-0152363
PR 28-JUN-1999;	99US-0140823	PR 10-SEP-1999;	99US-0153070
PR 29-JUN-1999;	99US-0140991	PR 13-SEP-1999;	99US-0153758
PR 30-JUN-1999;	99US-0141287	PR 15-SEP-1999;	99US-0154018
PR 01-JUL-1999;	99US-0141842	PR 16-SEP-1999;	99US-0154039
PR 02-JUL-1999;	99US-0142154	PR 20-SEP-1999;	99US-0154779
PR 06-JUL-1999;	99US-0142055	PR 22-SEP-1999;	99US-0155139
PR 08-JUL-1999;	99US-0142803	PR 23-SEP-1999;	99US-0155486
PR 09-JUL-1999;	99US-0142820	PR 24-SEP-1999;	99US-0156559
PR 12-JUL-1999;	99US-0142977	PR 28-SEP-1999;	99US-0156596
PR 13-JUL-1999;	99US-0143624	PR 29-SEP-1999;	99US-0157117
PR 14-JUL-1999;	99US-0143624	PR 05-OCT-1999;	99US-0157753
PR 15-JUL-1999;	99US-0144005	PR 06-OCT-1999;	99US-0157865
PR 16-JUL-1999;	99US-0144085	PR 07-OCT-1999;	99US-0158029
PR 16-JUL-1999;	99US-0144086	PR 08-OCT-1999;	99US-0158232
PR 19-JUL-1999;	99US-0144325	PR 12-OCT-1999;	99US-0158369
PR 19-JUL-1999;	99US-0144331	PR 13-OCT-1999;	99US-0159293
PR 19-JUL-1999;	99US-0144332	PR 13-OCT-1999;	99US-0159294
PR 19-JUL-1999;	99US-0144333	PR 13-OCT-1999;	99US-0159295
PR 19-JUL-1999;	99US-0144334	PR 14-OCT-1999;	99US-0159329
PR 19-JUL-1999;	99US-0144335	PR 14-OCT-1999;	99US-0159330
PR 20-JUL-1999;	99US-0144632	PR 14-OCT-1999;	99US-0159331
PR 20-JUL-1999;	99US-0144632	PR 14-OCT-1999;	99US-0159637
PR 20-JUL-1999;	99US-0144632	PR 14-OCT-1999;	99US-0159638
PR 21-JUL-1999;	99US-0144814	PR 18-OCT-1999;	99US-0159584
PR 21-JUL-1999;	99US-0145086	PR 21-OCT-1999;	99US-0160741
PR 21-JUL-1999;	99US-0145086	PR 21-OCT-1999;	99US-0160767
PR 22-JUL-1999;	99US-0145085	PR 21-OCT-1999;	99US-0160768

PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x AAC52409 ..

Align seg 1/1 to: AAC52409 from: 1 to: 696

26 Glutlleglyllylsyglu 32  
 ||||||||||||||||||  
 189 GAGATTGGGATCAAAAAGAG 209

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF11881

seq\_documentation\_block:

ID AAF11881 standard; cDNA; 710 BP.

XX AAF11881;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:4404.

XX Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.

XX WO200056762-A2.

PD 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

XX substrate of expressed sequence tags -

XX Claim 88; Page 1883-1884; 3161pp; English.

XX The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from  
 CC Fusarium venenatum; AAF1248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.

XX Sequence 710 BP; 135 A; 195 C; 148 G; 231 T; 1 other;

alignment\_scores:

Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x AAF11881/rev ..

Align seg 1/1 to reverse of: AAF11881 from: 1 to: 710

31 LysGluAlaGlyGluLysArg 37  
 ||||||||||||||||||

233 AAAAGAGCGGTCGAAAAGAG 213

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI97871

seq\_documentation\_block:

ID AAI97871 standard; cDNA; 721 BP.

XX AAI97871;

DT 13-NOV-2001 (first entry)

DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3946.

XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.

OS Homo sapiens.

XX WO20016719-A1.

PD 13-SEP-2001.

XX 02-MAR-2001; 2001WO-JP01629.

XX 07-MAR-2000; 2000JP-0159195.

XX (CHIB-) CHIBA PREFECTURE.

PA (HISM) HISAMITSU PHARM CO LTD.

PI Nakagawara A;

XX WPI; 2001-565584/63.

PT Nucleic acids originating in gene expressed in human neuroblastoma,

PT useful as probe or primer in diagnosing prognosis of human

PT neuroblastoma, malignancy and susceptibility indicator or tumour marker



PT for anti-cancer agents  
 XX  
 PS Claim 1; Page 2902; 2979pp; Japanese.  
 XX  
 CC The invention relates to novel genes (AA193926-AA197863) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and trkA genes.  
 XX  
 SQ Sequence 721 BP; 151 A; 200 C; 176 G; 190 T; 4 other;  
 alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-14 x AA197871/rev ..  
 Align seg 1/1 to reverse of: AA197871 from: 1 to: 721  
 32 GUAAGAGLGLTSAATGGLU 38  
 ||||||||||||||||||  
 434 GAGGCTGGGAGCAAGACGGA 414  
 seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAH77229  
 seq\_documentation\_block:  
 ID AAH77229 standard; DNA; 729 BP.  
 XX  
 AC AAH77229;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE DNA of atRSP41 5'-UTR.  
 XX  
 KW atRSP41 5'-UTR; thale cress; transgenic plant; virus resistance; ds;  
 KW altered reproductive capability; insect resistance; disease resistance;  
 KW Arabidopsis thaliana arginine/serine-type plant splicing factor; atRSP;  
 KW herbicide tolerance; antibiotic resistance.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT 5'UTR 1..729  
 FT /+tag= a  
 FT Intron 83..724  
 FT /\*tag= b  
 FT /number= "1"  
 XX  
 FT WO200175129-A1.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 02-APR-2001; 2001WO-BP03735.  
 XX  
 PR 03-APR-2000; 2000GB-0008120.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Barta A, Lopato S, Kalyna M:  
 XX  
 DR WPI; 2002-010793/01.  
 XX  
 XX DNA sequences comprising atRSP gene promoters, useful for expressing a  
 PT nucleotide sequence of interest for conferring a desirable phenotypic  
 PT trait to a plant, e.g. virus, insect or disease resistance, are new.  
 PS Claim 5; Page 30; 34pp; English.  
 XX

CC This polynucleotide sequence represents the DNA of the atRSP41 5'-UTR.  
 CC The invention relates to novel DNA sequences that function as promoters  
 CC of transcription of associated nucleotide sequences in plants. The  
 CC invention provides a DNA sequence capable of driving expression of an  
 CC associated nucleotide sequence, wherein said DNA sequence is obtainable  
 CC from genes of the atRSP (Arabidopsis thaliana arginine/serine-type plant  
 CC splicing factor) gene family. The invention further provides recombinant  
 CC DNA molecules comprising an atRSP41 promoter, the invention also provides  
 CC expression vectors and transgenic plants comprising the promoters of the  
 CC invention. The DNA sequence of the invention is useful for expressing a  
 CC nucleotide sequence of interest. The nucleotide sequences encode proteins  
 CC for conferring a desirable phenotypic trait to a plant transformed with  
 CC the protein, where such traits include antibiotic resistance, virus  
 CC resistance, insect resistance, disease resistance, or resistance to other  
 CC pests, herbicide tolerance, improved nutritional value, improved  
 CC performance in an industrial process or altered reproductive capability.  
 XX  
 SQ Sequence 729 BP; 168 A; 90 C; 137 G; 334 T; 0 other;  
 alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-14 x AAH77229/rev ..  
 Align seg 1/1 to reverse of: AAH77229 from: 1 to: 729  
 35 GLUYSARGLUHRLYSGLU 41  
 ||||||||||||||||||  
 145 GAGAAGAGAGAGACCAAGAC 125



OM of: US-08-973-363-14 to: EST: \* out\_format: pfs  
Date: Aug 3, 2002 5:46 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=framer-p2n: model -DEV=nlh  
-Q/cgnt\_1/USP10.spool/US08973363/runat\_01082002\_080123\_1814/app\_query.fasta\_1.638  
-DB=EST -QPM=fastap -SUFFIX=olip2n.rst -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-GAPOP=6.500 -CGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALLION=200 -THR\_SCORE=quality  
-THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US08973363 -CGNT\_1\_2938 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMECUT=120 -WARN\_TIMECUT=30 -NO\_XLPHY -WAIT -THREADS=1

## Search information block:

Query: US-08-973-363-14  
Query length: 45  
Database: EST: \*  
Database sequences: 13736207  
Database length: -1841657050  
Search time (sec): 4311.510000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
gb_est2:BG309136	+	9.00	130.42	39.34	911	I BG309136 HVSMC0002003f Hordeum
gb_est2:BG283942	+	9.00	148.38	51.08	1227	I BG283942 602408112F1 NIH.MGC.9
gb_est2:BI010841	-	8.00	143.61	94.26	198	I BI010841 MR2-EN0092-110101-006
gb_est2:BE787833	+	8.00	141.98	116.05	251	I BE787833 601482830P1 NIH.MGC.68
gb_est1:BB416876	+	8.00	141.64	121.30	264	I BB416876 BB416876 RIKEN full-16
gb_est1:AL377519	+	8.00	141.51	123.31	269	I AL377519 MMB3280981 MEBB Medic
gb_est2:N82639	+	8.00	140.88	133.70	295	I N82639 TGSTSTY24H01.r1 TGRH TdC
gb_est2:D71396	+	8.00	140.70	136.88	300	I D71396 CEK0744PFR yuji Kohara U
gb_est1:AM507143	+	8.00	140.43	141.62	303	I AM507143 EST00470 Plasmid Subst
gb_est2:BG664006	+	8.00	140.05	148.67	335	I BG664006 EST003096 Plasmidum Y
gb_est2:D73021	-	8.00	139.03	156.87	334	I D73021 CELN1126GR fuji Kohara U
gb_gss:BH582350	-	8.00	139.04	156.87	334	I BH582350 HOHMT1LTR BOMB Brassic
gb_est1:AA715354	-	8.00	138.92	159.24	386	I AA715354 HV36D06.r1 NCL.CGAP.BT
gb_est1:AA403024	-	8.00	138.42	163.38	423	I AA403024 UT-FR-BK0-aal-c-04-0-0
gb_est1:AA728877	+	8.00	138.13	150.20	441	I AA728877 hv37c09.r1 NCL.CGAP.BT
gb_est1:AM056878	+	8.00	137.71	199.25	465	I AM056878 ga07a06.r1 NIA Mouse N
gb_gss:AL257993	+	8.00	137.11	200.75	469	I AL257993 RPT1-23-120M12.TV RPT1
gb_est1:AL368855	+	8.00	137.65	202.25	473	I AL368855 MMB307D01P1 MEBB Medic
gb_est1:BB830260	+	8.00	137.65	202.25	473	I BB830260 MMB307D01P1 MEBB Medic
gb_est1:AL368855	+	8.00	137.65	202.25	473	I AL368855 MMB307D01P1 MEBB Medic
gb_gss:AL252252	+	8.00	137.15	215.68	509	I BH252252 SALK_013010 Arabidopsi
gb_gss:AL494245	+	8.00	137.06	218.28	518	I AL494245 IM0329P24F Mouse 10kb
gb_est1:AL367343	+	8.00	136.95	221.02	524	I AL367343 MEB414C12P1 MEBB Medic
gb_est1:AL367343	+	8.00	136.95	221.02	524	I AL367343 MEB414C12P1 MEBB Medic
gb_est1:AM171777	+	8.00	136.77	226.42	538	I AM171777 NIH00671E rootphos(-) M
gb_gss:AC0336204	+	8.00	136.03	249.13	600	I AC0336204 HS_5019.B1.P05.T7 RPT1
gb_gss:AC094721	+	8.00	135.74	256.57	626	I AC094721 Sheared DNA-52220.TF S
gb_est2:BT283828	-	8.00	135.15	278.74	629	I BT283828 EST47879 Rat Gene Ind
gb_gss:BH490490	-	8.00	135.14	278.74	629	I BH490490 BOMH1LTR BOMB Brassic
gb_est1:BB513748	+	8.00	135.00	281.18	683	I BB513748 BB513748 RIKEN full-16
gb_gss:BH531175	-	8.00	134.96	285.18	700	I BH531175 BOCGS09TF BOCB Brassic
gb_gss:BH465190	-	8.00	134.71	295.16	701	I BH465190 BOCGX09TF BOCB Brassic
gb_gss:BB662145	+	8.00	134.57	300.48	743	I BB662145 BB662145 RIKEN full-16
gb_gss:AC043398	+	8.00	134.43	306.15	759	I AC043398 Pan troglodytes DNA, c
gb_est2:BG455863	+	8.00	134.43	306.15	759	I BG455863 602540255P1 NIH.MGC.59

gb_gss:BH595299	-	8.00	134.32	310.03	770	I BH595299 BOCG528P BOCB Bras
gb_gss:AO896245	-	8.00	134.25	312.85	778	I AO896245 HS.3176.B1.G05.MR C
gb_gss:BH573931	-	8.00	134.25	312.85	778	I BH573931 BOCGV60TR BOCB Bras
gb_gss:BH581788	+	8.00	134.19	315.32	785	I BH581788 BOCGX88P BOCB Bras
gb_gss:BH503243	+	8.00	134.17	316.02	787	I BH503243 BOCXV46P BOCB Bras
gb_est2:BT490371	+	8.00	134.16	316.72	789	I BT490371 603031840A1 NIH.MGC
gb_gss:AG077821	+	8.00	134.15	317.08	790	I AG077821 Pan troglodytes DNA

seq\_name: gb\_est2:BG309136

seq\_documentation\_block:

LOCUS BG309136 911 bp mRNA linear EST 22-OCT-2001  
DEFINITION HVSMC0002003f Hordeum vulgare seedling shoot EST library  
HYDENA0003 (Etiolated and unstressed) Hordeum vulgare cDNA clone  
HVSMC0002003f, mRNA sequence.

ACCESSION BG309136  
VERSION BG309136.2 GI:16313822  
KEYWORDS EST.  
SOURCE barley.  
ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
I (bases 1 to 911)

REFERENCE Wing,R., Close,T.J., Kleinhoft,A., Wise,R., Begum,D., Frisch,D., Yu  
Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton  
,R.D., Oates,R. and Main,D.  
Development of a genetically and physically anchored EST resource  
for barley genomics: Morex unstressed seedling shoot cDNA library  
Unpublished (2001)  
On Feb 22, 2001 this sequence version replaced gi:13109983.  
Contact: Wing RA  
Clemson University Genomics Institute  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total bp bases = 143  
Seq primer: AATTAACTTCACAAAGG  
High quality sequence stop: 578.  
Location/Qualifiers

## FEATURES

1..911  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/cds\_xref="taxon:4513"  
/clone="HVSMC0002003f"  
/clone\_lib="Hordeum vulgare seedling shoot EST library  
HVSMC0003 (Etiolated and unstressed)"  
/tissue\_type="Seedling shoot"  
/lab\_host="TJC121"  
/note="Vector: lambdaZAP, Site:1: EcoRI; Site:2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, mystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedling shoots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give plasmid  
SK(-) cDNA phagemids. These steps were performed in the TJC  
Riverside laboratory at the University of California,  
Riverside (Choi, Close, Penlon). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates  
, Rambo, Main). The sequence has been trimmed to remove  
vector sequence and contains a minimum of 100 bases of  
phred value 20 or above. For more details on library  
preparation and sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhoft A, Wise R (2001)

Genetically and physically anchored EST resources for  
Barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/gppages/bgn/31/cover.html)"

BASE COUNT 257 a 216 c 345 g 93 t

ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x BG309136 ..

Align seg 1/1 to: BG309136 from: 1 to: 911

34 GtlylgsArgGltThrLysGltLys 42  
|||||  
827 GCGGAAAAGAGAGAAAGAGAGAG 853

seq\_name: gb\_est2:BG283942

seq\_documentation\_block:

LOCUS BG283942 1227 bp mRNA linear EST 21-FEB-2001  
DEFINITION 602408112F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4520114 5',  
mRNA sequence.

ACCESSION BG283942  
VERSION BG283942.1 GI:13034392

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1227)

NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov  
Plate: LLAM10417 row: h column: 03

High quality sequence stop: 168.

FEATURES

source

1..1227  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4520114"  
/clone\_lib="NIH\_MGC\_91"  
/tissue\_type="adenoecarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally: oligo-dT primed.  
Average insert size 1.4 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 587 a 208 c 388 g 44 t

ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x BG283942 ..

Align seg 1/1 to: BG283942 from: 1 to: 1227

34 GtlylgsArgGltThrLysGltLys 42  
|||||  
629 GCGGAAAAGAGAGAGAGAGAGAG 655

seq\_name: gb\_est2:BI010841

seq\_documentation\_block:

LOCUS BI010841 198 bp mRNA linear EST 13-JUN-2001  
DEFINITION MR2-EN0092-110101-006-d03\_1 EN0092 Homo sapiens cDNA, mRNA  
sequence.

ACCESSION BI010841  
VERSION BI010841.1 GI:14414912

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 198)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&ct=MR2-EN0092-  
110101-006-d03-1&f3=2001-01-11&f4=1)

Seq primer: puc 18 forward  
High quality sequence stop: 56.

Location/Qualifiers  
1..198  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="EN0092"  
/dev\_stage="Adult"

/note="Organ: lung,normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 38 a 51 c 31 g 78 t

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x BI010841/rev ..

Align seg 1/1 to reverse of: BI010841 from: 1 to: 198

35 GtlylgsArgGltThrLysGltLys 42

```

seq_name: gb_est2:BE787833                                     seq.documentat|on_block:                251 bp      mRNA       linear   EST 20-OCT-2000
LOCUS              BE787833                                      (bases 1 to 251)                      Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
DEFINITION         601482830F1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3885463 5',
                    MNNA sequence.
ACCESSION          BE787833
VERSION            BE787833
KEYWORDS           EST
SOURCE             human.
ORGANISM           homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteleostomi; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE          NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE              Unpublished (1999)
JOURNAL            Contact: Robert Strausberg, Ph.D.
COMMENT            Email: cgapbs-remail.nih.gov
                   Tissue Procurement: DCD/DTP/Gazdar
                   CDNA Library Preparation: Life Technologies, Inc.
                   DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN).
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LNLN at:
                   http://image.llnl.gov
                   Plate: LLM9660 row: 1 column: 08
                   High quality sequence stop: 188.
FEATURES           location/Qualifiers
source             1..251
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3885463"
                  /clone_1lb="NIH-MGC_68"
                  /library_type="large-cell carcinoma"
                  /lab_host="DHIOB (phage-resistant)"
                  /note="Organ: lung Vector: pCMV-Sport6 Site:1: Notf:
                  Site:2: SalI: Cloned unidirectionally.
                  Average insert size 1.8 kb. Library constructed by Life
                  Technologies."
BASE COUNT        93 a               58 c               74 g               26 t
ORIGIN
alignment_scores:
    Quality:      8.00                       length:      8
    Ratio:        1.000                     Gaps:          0
    Percent Simlarity: 100.000                 Percent Identity: 100.000
alignment_block:
US-08-973-363-14 x BE787833 ..
Align seg 1/1 to: BE787833 from: 1 to: 251
16 GlutylsgluAengluglutysasp 23
|||||||.....|
213 GAATAAGCAATGAGAAAGC 236
seq_name: gb_est1:BB416876
seq.documentat|on_block:
LOCUS              BB416876                                      264 bp      mRNA       linear   EST 16-JUL-2000
DEFINITION         CDNA clone C430048407 3', mRNa sequence.
ACCESSION          BB416876
VERSION            BB416876.1 GI:9238231
KEYWORDS           EST
SOURCE             house mouse.
ORGANISM           Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE	Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus
AUTHORS	1 (pages 1 to 264) Kono,H., Aizawa,K., Akhira,S., Akiyama,J., Arakawa,T., Carninci P., Endo,T., Fukuda,S., Fukunishi,Y., Hata,A., Hayatsu,N., Hirozane,T., Horii,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kituchi,N., Kiyosawa,H., Kojima,Y., Kono,S., Koya,S., Kuribara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Saki,C., Saito,K., Shibata,K., Shibata Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomiunga,N., Toyota P., Tsunoda,Y., Watanabe,K., Watanabe,S., Yamanura,T., Yamanaka,I., Yano,K., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino M., Yutanietsu,M., and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Kono,H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Toshinhide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Sueniwa-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagakawa,S., Sasaki ,N., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y. Thermosensitization and thermoactivation of chemolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95(2), 520-524 (1998) Itoh,M., Kitsuai,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki Y., and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.rtc.riken.go.jp">http://genome.rtc.riken.go.jp</a> ) for further details
FEATURES	Location/Qualifiers 1..264 /organism="Mus musculus" /db_xref="taxon:10090" /clone_id="C430048M07" /clone_1lb="RIKEN full-length enriched, 7 days embryo" /dev_stage="7 days embryo" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGACGATCCAGACGCTCTTTTCTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATCTTCGACGCTTAATTAATTAATTCGCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a pBluescript KS(+) after bulk excision from Lambda FIC I."
BASE COUNT	45 a 87 c 36 g 96 t
ORIGIN	
alignment_scores:	Quality: 8.00 Length: 8
	Ratio: 1.000 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 100.000
alignment_block:	

US-08-973-363-14 x BB416876/rev ..

Align seg 1/1 to reverse of: BB416876 from: 1 to: 264

30 LysylsGluAlaGlyGluLysArg 37  
 |||  
 158 AAAAAAGAGCGAGGAAAAAGG 135

seq\_name: gb\_est1:AL377519

seq\_documentation\_block:

LOCUS AL377519 269 bp mRNA linear EST 03-AUG-2000  
 DEFINITION MCB32B09R1 MCBB Medicago truncatula cDNA clone MCB32B09 T7, mRNA  
 sequence.

ACCESSION AL377519

VERSION AL377519.1 GI:9677271

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

Medicago.

1 (bases 1 to 269)

Journet,E.P., Crespeau,H., van Tuinen,D., Gouzy,J., Jaillon,O.,

Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson

,V. and Gamas,P.

Medicago truncatula ESTs from Sinorhizobium meliloti-induced root

nodules

Unpublished (2000)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de

Biologie Moleculaire des Relations Plantes-Microorganismes,

CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

Mc-est@toulouse.inra.fr Website :

http://sequence.toulouse.inra.fr/Mtruncatula.html).

Location/Qualifiers

1..269

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db\_xref="taxon:3880"

/clone\_lib="MCB32B09"

/clone\_lib="MCBB"

/tissue\_type="symbiotic root nodules"

/dev\_stage="harvested 4 days post inoculation with

Sinorhizobium meliloti"

/note="Vector: pBluescript PSK; Site\_1: EcoRI; Site\_2:

XhoI; Plants were grown in an aeroponic chamber on

nitrogen-rich medium for 21 days. Three days before

inoculation with Sinorhizobium meliloti, the medium was

replaced by N-free medium. Root nodules (+ short adjacent

root segments) were harvested 4 days post inoculation.

cDNA was prepared from polyA+ enriched RNA. The cDNA was

directionally ligated into uni-zap XR vector from

stratagene and packaged using Gigapack Gold packaging

extracts. Plasmids containing cDNA inserts were

mass-excised from phage stocks using EXASit helper phage

and propagated in SOLR cells. Clone ordering and

sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France)."

Sequence (Genoscope, Evry, France)."

BASE COUNT 104 a 38 c 43 g 84 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

Align seg 1/1 to: N82699 from: 1 to: 295

31 LysGluAsnGluGluLysArgGlu 38

|||

15 AAGGAGCGAGGAGGAGGAGGAG 38

US-08-973-363-14 x AL377519 ..

Align seg 1/1 to: AL377519 from: 1 to: 269

17 LysGluAsnGluGluLysArgGlu 24  
 |||  
 70 AAGGAGCGAGGAGGAGGAGGAG 93

seq\_name: gb\_est2:N82699

seq\_documentation\_block:

LOCUS N82699 295 bp mRNA linear EST 22-MAY-2000  
 DEFINITION TGESty24h01.r1 TGRH Tachyzoite cDNA Toxoplasma gondii cDNA clone  
 tgzy24h01.r1 5', mRNA sequence.

ACCESSION N82699

VERSION N82699.1 GI:1258452

KEYWORDS EST.

SOURCE Toxoplasma gondii.

ORGANISM Toxoplasma gondii

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;

Sarcocystidae; Toxoplasma.

1 (bases 1 to 295)

Hehl,A., Manger,I., Maria,M., Sibley,L.D., Ajioke,J.A., Aslett,M.A.,

Westerston,R.H. and Boothroyd,J.

Washu-Merck-Stanford-NIH Toxoplasma EST project

Unpublished (1996)

Contact: Maria M

Washu-Merck EST project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: toxowatson.wustl.edu

David Sibley at toxowatson.wustl.edu for further information

relating to organism, clone or library availability.

Seq primer: T3

High quality sequence stop: 283.

Location/Qualifiers

1..295

/organism="Toxoplasma gondii"

/strain="RH"

/db\_xref="taxon:5811"

/clone\_lib="tgzy24h01.r1"

/clone\_lib="TGRH Tachyzoite cDNA"

/lab\_host="XLI-Blue MRP"

/note="Vector: lambda ZAP; Site\_1: EcoRI; Site\_2: XhoI;

Toxoplasma RH strain tachyzoites were grown in human

foreskin fibroblast cultures in vitro. The library was

constructed by K.L. Wan, Cambridge University. cDNAs were

synthesized from polyA RNAs by oligo d(T) priming and

directionally cloned into the EcoRI to XhoI sites of the

lambda ZapII vector using the ZAP-cDNA synthesis kit

(Stratagene). WARNING: the library contains a small

percentage of cDNAs derived from the human host cells."

BASE COUNT 85 a 71 c 109 g 27 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x N82699 ..

Align seg 1/1 to: N82699 from: 1 to: 295

31 LysGluAsnGluGluLysArgGlu 38

|||

15 AAGGAGCGAGGAGGAGGAGGAG 38

```

seq_name: gb_est2:D71396
seq_documentation_block:
  LOCUS D71396 300 bp mRNA linear EST 18-OCT-1999
  DEFINITION CEK074FR yuji kohara unpublished cDNA:Strain N2 hermaphrodite
  EMBL074 Caenorhabditis elegans cDNA clone YK747 3', mRNA sequence.
  ACCESSION D71396
  VERSION D71396.1 GI:11111103
  KEYWORDS EST.
  SOURCE Caenorhabditis elegans.
  ORGANISM Caenorhabditis elegans.
  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae
  ; Rhabditidae; Pelodierinae; Caenorhabditis.
  1 (bases 1 to 300)
  Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
  Tabara,H.
  TITLE Toward an expression map of the C.elegans genome
  JOURNAL Unpublished (1994)
  COMMENT Contact: Yuji Kohara
  Genome Biology Lab.
  National Institute of Genetics
  Yata 1111 Mishima, Shizuoka 411, Japan
  Tel: 81-559-81-6854
  Fax: 81-559-81-6853
  Email: ykoha@lab.nig.ac.jp.
  Location/Qualifiers
  1..300
  /organism="Caenorhabditis elegans"
  /strain="N2"
  /db_xref="taxon:6239"
  /clone="YK747"
  /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
  hermaphrodite embryo"
  /sex="hermaphrodite"
  /dev_stage="embryo"

BASE COUNT 122 a 45 c 67 g 58 t 8 others
ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
  US-08-973-363-14 x D71396 ..
  Align seg 1/1 to: D71396 from: 1 to: 300

35 GlutylsArgLutThrLysGluLys 42
  ||||||||||||||||||
  20 GAAAAAAGCGAAACGAAAGAAAA 43

seq_name: gb_est1:AM507143
seq_documentation_block:
  LOCUS AM507143 303 bp mRNA linear EST 01-JUN-2000
  DEFINITION EST00470 Plasmid Subtractive Library of Rat Cerebrum (stroke)
  Rattus norvegicus cDNA clone 3rdF2, mRNA sequence.
  ACCESSION AM507143
  VERSION AM507143.1 GI:8133308
  KEYWORDS EST.
  SOURCE Norway rat.
  ORGANISM Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 303)
  Zhao,B., Wang,X.M., Zhang,Q., Zhu,Z.M., Zhu,S.J. and Hui,R.T.
  TITLE Cloning of Genes Responsible for Stroke
  JOURNAL Unpublished (2000)
  COMMENT Contact: Bin Zhao
  Molecular Medical Center for Cardiovascular Disease

```

```

Cardiovascular Institute, CAMS & PUMC
167, Bei Li Shi Lu, 100037, Beijing, P.R. China
Tel: 86-10-68314466 ext 8136
Fax: 86-10-68313390
Email: binzhao@ctonghua.com.cn
Seq primer: M13R
POLYA-No.

FEATURES
  source
  1..303
  /organism="Rattus norvegicus"
  /strain="Wistar"
  /db_xref="taxon:10116"
  /clone="3rdF2"
  /clone_lib="Plasmid Subtractive Library of Rat Cerebrum
  (stroke)"
  /sex="male"
  /tissue_type="Cerebrum"
  /dev_stage="8-12 weeks old"
  /note="Vector: pMD-19 (Takara, D504 CA). The rat stroke
  model was developed by Wang Xianmei,Zhu Zhiming, Zhu
  Shanjun et al at 3rd Military Medical University Chongqing
  , China. The Suppression Subtraction Hybridization (SSH)
  was conducted at Fu Wai Hospital by Zhao Bin and Wang
  Xianmei in Beijing, China. Message RNA was extracted from
  pooled whole cerebrum from stroke and control rats
  respectively. SSH had been conducted following
  manufacture's manual using 5ug mRNA. The SSH library was
  constructed by Zhao Bin and Wang Xianmei using pMD18-T
  vector. Randomly picked clones were sequenced at ABI 377
  using M13R primer.(AGCGATTAACATTCACACAG) by Wang Xianmei
  , Zhao Bin and Zhang Qian."

BASE COUNT 90 a 63 c 60 g 90 t
ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

Alignment_block:
  US-08-973-363-14 x AM507143 ..
  Align seg 1/1 to: AM507143 from: 1 to: 303

11 LysLysIleLysAlaGluLysGlu 18
  ||||||||||||||||||
  209 AAGAAATTAAAGCTGAGAAAGAG 232

seq_name: gb_est2:BG604006
seq_documentation_block:
  LOCUS BG604006 315 bp mRNA linear EST 14-AUG-2001
  DEFINITION EST503096 Plasmodium yoelii sporozoite cDNA Plasmodium yoelii cDNA
  clone PYCDX37, mRNA sequence.
  ACCESSION BG604006
  VERSION BG604006.1 GI:15154020
  KEYWORDS EST.
  SOURCE Plasmodium yoelii.
  ORGANISM Plasmodium yoelii
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
  1 (bases 1 to 315)
  Kappe,S.H.I., Gardner,M.J., Brown,S.M., Ross,J., Matuschewski,K.,
  Ribeiro,U.M., Adams,J.H., Quackenbush,J., Cho,J., Carucci,D.J.,
  Hoffman,S.L. and Nussenzweig,V.
  TITLE Exploring the transcriptome of the malaria sporozoite stage
  JOURNAL Proc. Natl. Acad. Sci. U. S. A. 98 (17), 9895-9900 (2001)
  MEDLINE 21396355
  COMMENT Contact: Malcolm J. Gardner
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301 838 3519

```

BASE COUNT  
ORIGIN

```
seq_name: gb_est1:AA/133
```

seq\_documentation\_block:



```
LOCUS       AA715354               386 bp    mRNA    linear    EST 22-JAN-1998
DEFINITION   nv36d06.r1 NCI_CGAP_Br5 Homo sapiens cDNA clone IMAGE:1222283, mRNA
sequence.
ACCESSION   AA715354
VERSION     AA715354.1   GI:2727628
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 386)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
COMMENT     Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
            Ph.D. student, Rodrigo F. Chuquil, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            CDNA Library Preparation: David B. Krizman, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert length: 604 Std Error: 0.00
            Seq primer: -28m3 rev1 ET from Amersham
            High quality sequence stop: 370.
FEATURES
    source
        1..386
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1222283"
            /clone_lib="NCI_CGAP_Br5"
            /sex="female"
            /tissue_type="infiltrating ductal carcinoma"
            /lab_host="DH10B"
            /note="Organ: breast; Vector: PAMP10; mRNA made from
            infiltrating ductal carcinoma, cDNA made by oligo-dr
            priming. Non-directionally cloned. Size-selected on
            agarose gel, average insert size 600 bp."
BASE COUNT   76 a      87 c      44 g      179 t
ORIGIN
alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:      0
    Percent Similarity: 100.000   Percent Identity: 100.000
Alignment_block:
    US-08-973-363-14 x AA715354/rev ..
Align seg 1/1 to reverse of: AA715354 from: 1 to: 386
    35 G1ULYARG1UTHTLysG1ULys 42
    ||||||||||||||||||||
    139 GAAAAAAGGAAACGAAAGAGAAA 116
seq_name: gb_est1:AA403024
seq_documentation_block:
    LOCUS       AA403024               391 bp    mRNA    linear    EST 16-FEB-2000
DEFINITION   UT-HF-BRC-aat-c-04-0-UT.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3053934 5', mRNA sequence.
ACCESSION   AA403024
VERSION     AA403024.1   GI:6921838
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 391)
AUTHORS     NIH-MGC http://imgc.ncbi.nlm.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Eco RI site shown at the beginning of the sequence.
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: M.B. Soares Lab
            CDNA Library Arrayed by: M.B. Soares Lab
            DNA Sequencing by: M.B. Soares Lab
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward.
FEATURES
    source
        1..391
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3053934"
            /clone_lib="NIH_MGC_36"
            /tissue_type="lymph"
            /cell_type="germinal center B cells"
            /cell_line="MGC85"
            /lab_host="DH10B (LTI)"
            /note="Vector: pT73-Pac; Site_1: NotI; Site_2: Eco RI;
            Constructed from size fractionated cytoplasmic mRNA
            (0.5-1.5kb). Directionally cloned. Cells provided by Louis
            M. Staudt, Ph.D. Library preparation by Maria de Fatima
            Bonafide, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT   96 a      132 c      95 g      68 t
ORIGIN
alignment_scores:
    Quality:      8.00      Length:      8
    Ratio:        1.000     Gaps:      0
    Percent Similarity: 100.000   Percent Identity: 100.000
Alignment_block:
    US-08-973-363-14 x AA403024 ..
Align seg 1/1 to: AA403024 from: 1 to: 391
    15 A1AG1ULYSG1UASng1ULys 22
    ||||||||||||||||||||
    7 GCCGAAAAAGAAATGAAAGAAAAA 30
seq_name: gb_est1:AA728877
seq_documentation_block:
    LOCUS       AA728877               423 bp    mRNA    linear    EST 22-JAN-1998
DEFINITION   nv37c09.r1 NCI_CGAP_Br5 Homo sapiens cDNA clone IMAGE:1222384, mRNA
sequence.
ACCESSION   AA728877
VERSION     AA728877.1   GI:2750236
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 423)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
COMMENT     Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
            Ph.D. student, Rodrigo F. Chuquil, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            CDNA Library Preparation: David B. Krizman, Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
```

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/db/brp/image/image.html](http://www.bio.lnl.gov/db/brp/image/image.html)  
 Insert length: 594 Std Error: 0.00  
 Seq primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 359.

FEATURES

source

Location/Qualifiers  
 1. 423  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1222384"  
 /clone\_lib="NCI-CGAP\_Br5"  
 /sex="female"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: breast; Vector: pAMP10; mRNA made from  
 infiltrating ductal carcinoma, cDNA made by oligo-dT  
 priming. Non-directionally cloned. Size-selected on  
 agarose gel, average insert size 600 bp."  
 BASE COUNT 88 a 95 c 46 g 194 t  
 ORIGIN

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-14 x AA728877/rev ..

Align seg 1/1 to reverse of: AA728877 from: 1 to: 423

35 GLULYSARGIUTRNLYSGULYS 42  
 ||||||||||||||||||||  
 139 GAAAAAGGGAACGAAAGAGAAA 116